TECHNISCHE
UNIVERSITÄT
DARMSTADT

FOUNDATIONS, INFERENCE, AND DECONVOLUTION IN IMAGE RESTORATION

Dissertation approved by
TECHNISCHE UNIVERSITÄT DARMSTADT
Fachbereich Informatik

for the degree of
Doctor rerum naturalium
(Dr. rer. nat.)

by

KEVIN SCHELTEN
Dipl.-Math.

Examiner: Prof. Stefan Roth, PhD
Co-examiner: Dr. Sebastian Nowozin

Darmstadt, 2017
D17

Kevin Schelten: *Foundations, Inference, and Deconvolution in Image Restoration*

LOCATION:
Darmstadt, Technische Universität Darmstadt

PUBLICATION YEAR ON TUPRINTS:
2018

DATE OF DEFENSE:
August 22, 2017

ABSTRACT

IMAGE restoration is a critical preprocessing step in computer vision, producing images with reduced noise, blur, and pixel defects. This enables precise higher-level reasoning as to the scene content in later stages of the vision pipeline (*e.g.*, object segmentation, detection, recognition, and tracking). Restoration techniques have found extensive usage in a broad range of applications from industry, medicine, astronomy, biology, and photography. The recovery of high-grade results requires models of the image degradation process, giving rise to a class of often heavily underconstrained, inverse problems. A further challenge specific to the problem of blur removal is noise amplification, which may cause strong distortion by ringing artifacts. This dissertation presents new insights and problem solving procedures for three areas of image restoration, namely *(1)* model foundations, *(2)* Bayesian inference for high-order Markov random fields (MRFs), and *(3)* blind image deblurring (deconvolution).

As basic research on model foundations, we contribute to reconciling the perceived differences between probabilistic MRFs on the one hand, and deterministic variational models on the other. To do so, we restrict the variational functional to locally supported finite elements (FE) and integrate over the domain. This yields a sum of terms depending locally on FE basis coefficients, and by identifying the latter with pixels, the terms resolve to MRF potential functions. In contrast with previous literature, we place special emphasis on robust regularizers used commonly in contemporary computer vision. Moreover, we draw samples from the derived models to further demonstrate the probabilistic connection.

Another focal issue is a class of high-order Field of Experts MRFs which are learned generatively from natural image data and yield best quantitative results under Bayesian estimation. This involves minimizing an integral expression, which has no closed form solution in general. However, the MRF class under study has Gaussian mixture potentials, permitting expansion by indicator variables as a technical measure. As approximate inference method, we study Gibbs sampling in the context of non-blind deblurring and obtain excellent results, yet at the cost of high computing effort. In reaction to this, we turn to the mean field algorithm, and show that it scales quadratically in the clique size for a standard restoration setting with linear degradation model. An empirical study of mean field over several restoration scenarios confirms advantageous properties with regard to both image quality and computational runtime.

This dissertation further examines the problem of blind deconvolution, beginning with *localized* blur from fast moving objects in the scene, or from camera defocus. Forgoing dedicated hardware or user labels, we rely only on the image as input and introduce a latent variable model to explain the non-uniform blur. The inference procedure estimates freely varying kernels and we demonstrate its generality by extensive experiments.

We further present a discriminative method for blind removal of camera shake. In particular, we interleave discriminative non-blind deconvolution steps with kernel estimation and leverage the error cancellation effects of the Regression Tree Field model to attain a deblurring process with tightly linked sequential stages.

ZUSAMMENFASSUNG

BILDWIEDERHERSTELLUNG ist ein entscheidender Vorverarbeitungsschritt im maschinellen Sehen (Computer Vision), welcher Bilder mit weniger Rauschen, Unschärfe und Pixelfehlern erzeugt. Dies ermöglicht es, präzise Folgerungen über den Bildinhalt auf einer höheren Abstraktionsebene in den späteren Stadien der Vision-Bearbeitungskette zu ziehen (z.B. Objektsegmentierung, -detektion, -erkennung und -verfolgung). Bildwiederherstellungstechniken haben umfangreiche Verwendung in einer breiten Auswahl von Anwendungen aus Industrie, Medizin, Astronomie, Biologie und Fotografie gefunden. Die Berechnung von hochwertigen Ergebnissen erfordert Modelle des Bildverschlechterungsprozesses, die zu einer Klasse von oft stark unterbestimmten, inversen Problemen führen. Eine weitere, für die Entfernung von Unschärfe spezifische Herausforderung, ist die Verstärkung des Rauschens, welche eine schwere Verzerrung durch Wellenartefakte verursachen kann. Die vorliegende Dissertation präsentiert neue Erkenntnisse und Problemlösungsverfahren für drei Bereiche der Bildwiederherstellung, nämlich *(1)* Modellgrundlagen, *(2)* Bayessche Inferenz für Markov Random Fields (MRFs) mit hoher Cliquengröße und *(3)* blinde Bildschärfung (Dekonvolution).

Als Basisforschung zu Modellgrundlagen tragen wir dazu bei, die Unterschiede in der Wahrnehmung von probabilistischen MRFs einerseits und deterministischen Variationsmodellen andererseits auszugleichen. Dazu beschränken wir das Variationsfunktional auf Finite Elemente (FE) mit lokalem Träger und integrieren über den Definitionsbereich. Dies ergibt eine Summe von Termen, die lokal von FE-Basiskoeffizienten abhängen, und indem wir die letzteren mit Pixeln gleichsetzen, lösen sich die Terme zu MRF-Potentialfunktionen auf. Im Gegensatz zur bisherigen Literatur legen wir eine besondere Betonung auf robuste Regularisierungsfunktionen, die in der zeitgenössischen Computer Vision üblich sind. Außerdem erzeugen wir Zufallsstichproben aus den abgeleiteten Modellen, um die probabilistische Verbindung weiter zu belegen.

Ein weiteres Schwerpunktthema ist eine Klasse von Field of Experts MRFs mit hoher Cliquengröße, welche generativ aus natürlichen Bilddaten gelernt sind und quantitativ beste Ergebnisse unter Bayes-Schätzung liefern. Dies beinhaltet die Minimierung eines Integralausdrucks, wozu i.A. keine geschlossene Lösung existiert. Allerdings hat die betrachtete Klasse von MRFs Gaussian Mixture Potentiale, welche eine Erweiterung durch Indikatorvariablen als technische Maßnahme zulassen. Als approximative Inferenzmethode untersuchen wir Gibbs-Sampling im Kontext nicht-blinder Dekonvolution

und erhalten exzellente Resultate, jedoch auf Kosten eines hohen Rechenaufwands. Als Reaktion darauf wenden wir uns dem Mean Field Algorithmus zu und zeigen, dass dieser quadratisch in der Cliquengröße skaliert, falls eine in der Bildwiederherstellung übliche Formulierung mit linearem Verschlechterungsmodell zu Grunde liegt. Eine empirische Studie der Mean Field Methode über mehrere Wiederherstellungsszenarien bestätigt die vorteilhaften Eigenschaften in Bezug auf sowohl Bildqualität als auch Rechenzeit.

In dieser Dissertation wird zudem das Problem der blinden Dekonvolution untersucht, beginnend mit *lokaler* Unschärfe, die durch schnell bewegte Objekte in der Szene, oder durch begrenzte Schärfentiefe verursacht wird. Dabei verzichten wir auf dedizierte Hardware oder Markierungen des Anwenders, verwenden nur das Bild als Eingabe und führen ein mit latenten Variablen versehenes Modell ein, um die nicht einheitliche Unschärfe zu erklären. Der Inferenzprozess schätzt frei variierende Faltungskerne, und wir belegen dessen allgemeine Anwendbarkeit durch umfangreiche Experimente.

Außerdem präsentieren wir eine diskriminative Methode zur blinden Entfernung von Verwacklungsunschärfe. Insbesondere verzahnen wir diskriminative, nicht-blinde Dekonvolutionsschritte mit Faltungskernschätzung und setzen die Fehler ausgleichende Wirkung des Regression Tree Field Modells ein, um einen Dekonvolutionsprozess mit eng verknüpften, sequentiellen Stufen zu gewinnen.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

BP	belief propagation (Pearl, 1982)
BSDS	Berkeley segmentation data set and benchmark (Arbelaez et al., 2011)
CCD	charge-coupled device
CG	conjugate gradient method (Hestenes and Stiefel, 1952)
CRF	conditional random field
CT	computed tomography
EFF	Efficient Filter Flow (Hirsch et al., 2010)
EM	expectation maximization (Dempster et al., 1977)
EPLL	Expected Patch Log Likelihood (Zoran and Weiss, 2011)
EP	expectation propagation (Minka, 2001)

FDM	finite difference method
FD	finite difference
FEM	finite element method
FE	finite element
FFT	fast Fourier transform
FOE	Field of Experts (Roth and Black, 2009)
FRAME	Filters, Random Fields and Maximum Entropy (Zhu et al., 1998)
GMM	Gaussian mixture model
GPU	graphics processing unit
GSM	Gaussian scale mixture
IRLS	iteratively reweighted least squares
ISO	international organization for standardization
KL	Kullback-Leibler divergence (Kullback and Leibler, 1951)
LBFGS	limited-memory Broyden–Fletcher–Goldfarb–Shanno
LED	light-emitting diode
MAE	mean absolute error
MAP	maximum a-posteriori
MCMC	Markov chain Monte Carlo
MMSE	minimum mean squared error
MRF	Markov random field
MRI	magnetic resonance imaging
PDE	partial differential equation
POE	Product of Experts (Hinton, 1999)
PSNR	peak signal-to-noise ratio
ROF	Rudin-Osher-Fatemi model (Rudin et al., 1992)
ROI	region of interest
RTF	Regression Tree Field (Jancsary et al., 2012a)
SSD	sum of squared differences

SSIM	structural similarity index (Wang et al., 2004)
TIFF	tagged image file format
TV	total variation
UAV	unmanned aerial vehicle
VOC	PASCAL visual object classes (Everingham et al., 2015)
dB	decibel

1 INTRODUCTION

CONTENTS

With 77% of Americans owning a smartphone in early 2017 (Pew Research Center, 2017), mobile phone cameras are ubiquitous. Their popularity lies partly rooted in being easily available to capture the spontaneous moment, while permanent connectivity paired with social networking applications allows to immediately share photos and engage with a broad online community. In fact, smartphones are the most popular cameras on the photo sharing application Flickr (Flickr, 2017). However, mobile phone imagery is often damaged by displeasing blur and noise artifacts. One reason is that mobile phones are frequently challenging to hold steady while tapping to focus and activating the shutter release. Additionally, the camera lens is easily smudged by fingerprints or dust, causing image details to smear and lose clarity.

Nevertheless, the more computers *"weave themselves into the fabric of everyday life"* (Weiser, 1991), the more processing power to recover from imaging errors is at our immediate disposal (*e.g.*, smartphones have adequate processors and memory for many applications). The rapid rise of cloud computing[1] also gives access to vast internet-based computing resources, which constantly scale to meet changing demands. Moreover, there is an unprecedented volume of data available over the internet for the training and evaluation of image enhancement algorithms. *E.g.*, in 2016, more than 1.5 million photos were uploaded to Flickr every day on average (Michel, 2017).

In a parallel development, many businesses and industries are becoming increasingly software-oriented, while cameras are manufactured as ever more sophisticated tools, *e.g.*, being 3D-printed to such a miniature scale they can be injected by a syringe (Gissibl et al., 2016). Against this technological backdrop, we specify the following, main topical focus of this dissertation.

1 Amazon cloud computing reports rises in sales of 55%, 70%, and 49% in 2016, 2015, and 2014 (Amazon.com, Inc., 2017).

1.1 PROBLEM STATEMENT

The goal of computer vision is to draw information from images automatically. As stated by Horn (1986), *"A machine vision system analyzes images and produces descriptions of what is imaged."* Unsurprisingly, the performance of such systems depends substantially on the quality of the input images: The sharper and clearer they are, the easier it is to make inferences regarding their content. For example, removing blur may significantly enhance face recognition (Nishiyama et al., 2011). Correspondingly, image restoration is the subdiscipline of computer vision concerned with recovering a clean image from one corrupted by noise, blur or defective pixels. This dissertation addresses several problems in modeling and inference for image restoration, with image deblurring[2] as a main underlying theme.

1.2 MOTIVATION

Digital image enhancement is of fundamental importance for industry and the sciences, while playing a central role in the art of photography. In the following, we present a diverse list of applications.

INDUSTRY. In photogrammetry as well as for the aerial inspection of essential civil infrastructures such as power lines, cooling towers, bridges, train tracks, and pipelines, the use of unmanned aerial vehicles (UAVs) can lead to blurred imagery due to the motion of the drone, especially in cases of sudden turbulences or high wind speeds. Additional processing to recover image details may be required (Hammer et al., 2007; Sieberth et al., 2014). Image enhancement is also relevant for quality control of moving parts on a conveyor belt, *e.g.*, in the context of automatically inspecting food products such as fruits, vegetables, and grains (Brosnan and Sun, 2004; Topalova and Tzokev, 2010). Another industrial area to benefit from modeling image degradation is visual tracking, where low-quality, motion blurred video frames occur frequently and need to be processed robustly (Bascle et al., 1996; Jin et al., 2005; Wu et al., 2011). The accuracy of biometric systems such as face or iris recognition can also be improved by careful consideration or outright removal of camera defocus (Kang and Park, 2007; Ahonen et al., 2008; Nishiyama et al., 2011).

MEDICINE. Medical data originating *e.g.* from magnetic resonance imaging (MRI), computed tomography (CT) scans or ultrasound imaging contain random noise arising from the respective physical procedures leading to image acquisition. As such random fluctuations

2 Throughout this thesis, the terms "deblurring" and "deconvolution" are used interchangeably.

adversely affect further processing steps for diagnostic analysis such as segmentation of salient image structures, noise removal is an essential technological component (Manjón et al., 2008; Li et al., 2014). In addition, we note that for MRI, object motion including the heart beat can cause images to be damaged by blur in a locally restricted area (Hirsch et al., 2010). To give another, more specific example, blind deblurring of CT scans permits to recover fine details of the temporal bone, which is the part of the human skull supporting the inner ear structure. This facilitates the preoperative planning of cochlear implants to counteract severe hearing impairment (Jiang et al., 2003). Moreover, liquid lens cameras have properties such as reduced volume and absence of moving parts which make them relevant to laparoscopic surgery, a minimally invasive technique requiring only small incisions for the insertion of surgical instruments including a fiber optic camera. Tzeng et al. (2010) enhance the resulting images, which are distorted by the chromatic aberration of unfocused red and blue color channels.

ASTRONOMY. Astronomical images are degraded by the Earth's atmospheric turbulence, which leads to refractive index fluctuations along the optical path and causes non-stationary blur in the sense that it changes temporally and spatially across the image plane. The established technique of *lucky imaging* consists of taking many images of very short exposures in an attempt to capture near-static atmospheric conditions and then combining the best images to one (Tubbs, 2003). Recent blind deblurring approaches for this task also demonstrate high performance (Harmeling et al., 2009; Hirsch et al., 2010; Zhu and Milanfar, 2013). Furthermore, the particular statistics of astronomical images motivate custom noise removal techniques (Burger et al., 2011; Beckouche et al., 2013).

BIOLOGY. Fluorescence microscopy is an important tool for biological analysis due to the detailed and accurate 3D images of cellular structures it produces. Thereby, the sample under study is stained with chemicals if it is not of naturally fluorescing material. Subsequently exciting the specimen by illumination causes it to emit light which is filtered and captured by the imaging system. Different focal calibrations of the microscope yield 2D sample slices situated at varying depths. However, a significant source of image degradation is the light emitted by out-of-focus planes, which causes sample-specific distortion. Specialized image restoration techniques nonetheless permit to recover intricate cellular details (Sarder and Nehorai, 2006; Kenig et al., 2010; Keuper et al., 2013).

PHOTOGRAPHY. Adams (1995a,b,c) places great emphasis on the craft and technical equipment involved in the art of photography. In

today's *digital darkroom* (Holzmann, 1988), computer algorithms form the centerpiece. Digital photography editing not only allows to enhance composition, contrast, brightness, hue, and saturation, but also includes image restoration functionality such as spot removal, noise suppression, and image sharpening. Support for camera shake reduction was introduced to Adobe Photoshop CC in 2013 (Wang, 2013). Today, many applications with similar functionality run directly on smartphones. One may further remark that post-processed photos are particularly widespread in social networking services such as Twitter, Instagram, and Flickr. In that regard, a recent study by Bakhshi et al. (2015) reveals that on Flickr, filtered images are 21% more likely to be viewed and 45% more likely to elicit interaction from other users in the form of commentary.

1.3 CHALLENGES

Having shed light on various applications of image restoration, we now clarify the main obstacles to overcome in this research area.

INVERSE PROBLEMS. Computer vision utilizes forward models to describe how the world is projected onto the camera plane, but its true nature is the inverse endeavor, *i.e.*, the description of the original scene from the observed pixel data. In the context of image restoration, a general forward model is

$$\mathbf{y} = \mathbf{K}\mathbf{x} + \mathbf{n}, \tag{1.1}$$

where $\mathbf{y} \in \mathbb{R}^n$ denotes the observed image, $\mathbf{K} \in \mathbb{R}^{n,m}$ is a linear matrix, $\mathbf{x} \in \mathbb{R}^m$ is the hidden image, and $\mathbf{n} \in \mathbb{R}^n$ is additive noise. Equation (1.1) encompasses the problems of denoising (where $\mathbf{K} = \mathbf{I}$, the identity matrix), deblurring (where $\mathbf{K}$ expresses convolution), and superresolution (where $\mathbf{K}$ expresses blur and downsampling). The difficulty of inverting Equation (1.1) to solve for $\mathbf{x}$ lies in missing information. In general, $m \geqslant n$, and the noise perturbation $\mathbf{n}$ is unknown. The matrix $\mathbf{K}$ may not be invertible and is unspecified in the cases of blind deblurring and superresolution (Michaeli and Irani, 2013). As Lanczos (2012) remarks, *"... a lack of information cannot be remedied by any mathematical trickery"*. We will see below that the common approach to circumvent this is to reformulate the problem.

ILL-POSEDNESS AND ILL-CONDITIONEDNESS. Let us consider the easier problem of non-blind image deconvolution, in which case the blur matrix $\mathbf{K}$ is known. If we further assume a noiseless system and uniform blur, *i.e.*, $\mathbf{K}\mathbf{x} = \mathbf{k} \otimes \mathbf{x}$ for some kernel $\mathbf{k}$, Equation (1.1) is equivalent to

$$\mathcal{F}(\mathbf{y}) = \mathcal{F}(\mathbf{k}) \cdot \mathcal{F}(\mathbf{x}) \tag{1.2}$$

in frequency domain, with $[\cdot]$ denoting element-wise product. We note that in case the kernel $\mathbf{k}$ has zero frequencies, which may occur for motion or defocus blur, Equation (1.2) is ill-posed in the sense that many solutions for the hidden image $\mathbf{x}$ exist. On the other hand, if the kernel is invertible, the presence of noise is a disturbing factor, since the noise contribution $\mathcal{F}(\mathbf{n})/\mathcal{F}(\mathbf{k})$ can heavily degrade the result, especially when the kernel has frequencies of very small magnitude. The ill-posedness of blind deblurring is perhaps best illustrated by the undesirable *no-blur* solution, which also occurs in practice when deblurring algorithms fail: Equation (1.2) is trivially fulfilled when the sought after blur $\mathbf{k}$ and sharp image $\mathbf{x}$ equal the identity kernel δ and blurred input $\mathbf{y}$, respectively.

MODELING AND INFERENCE. It is common to mitigate inverse problems by determining approximate solutions. We will give an instructive example for the case of non-blind deblurring in the beginning of this paragraph. A first step is to relax the equality constraint of Equation (1.2) by a least-squares cost term, namely $\|\mathbf{y} - \mathbf{K}\mathbf{x}\|^2$. To control noise amplification upon kernel inversion, we further impose quadratic penalties $\|\nabla\mathbf{x}\|^2 = \|\mathbf{f}_x \otimes \mathbf{x}\|^2 + \|\mathbf{f}_y \otimes \mathbf{x}\|^2$ on the image gradients, with $\mathbf{f}_x$ and $\mathbf{f}_y$ denoting derivative filters. In summary, the objective function to optimize for the hidden image $\mathbf{x}$ is $\|\mathbf{y} - \mathbf{K}\mathbf{x}\|^2 + \lambda\|\nabla\mathbf{x}\|^2$, where the parameter λ modulates the influence of data and smoothness terms on the result, allowing to adapt to image noise of differing magnitude. In the case of a uniform blur $\mathbf{k}$, it is not hard to verify that the cost function minimum is specified in closed form by the expression

$$\mathbf{x} = \mathcal{F}^{-1}\left(\frac{\overline{\mathcal{F}(\mathbf{k})}\cdot\mathcal{F}(\mathbf{y})}{|\mathcal{F}(\mathbf{k})|^2 + \lambda\left(|\mathcal{F}(\mathbf{f}_x)|^2 + |\mathcal{F}(\mathbf{f}_y)|^2\right)}\right), \tag{1.3}$$

which follows the *Wiener filter* paradigm (Wiener, 1964; Levin et al., 2009; Xu and Jia, 2010)[3] . Let us now illustrate the modeling challenges faced by contemporary research, contrasting with elements of the traditional derivation above where feasible.

Image Formation. Implicit in Equations (1.2) and (1.3) is the assumption of spatially invariant blur, *i.e.*, the same kernel acts on every site of the image plane. However, in the case of camera shake, in-plane rotation causes the blur to vary spatially. A more accurate forward model is to describe the blurred image as an aggregation of latent image homographies expressing a discrete range of camera poses and weighted by exposure time (Whyte et al., 2010; Gupta et al., 2010; Tai et al., 2011), although we note that uniform deblurring nonetheless demonstrates high benchmark performance (Köhler

3 Note that Equation (1.3) uses element-wise multiplication, division, and squared modulus notation.

et al., 2012; Xu et al., 2013). Furthermore, the type of *localized blur* caused by rapid object motion in the scene or camera defocus is also incompatible with the invariant convolution model. Spatially limited blur estimation procedures are needed in this case. In particular, it is essential to distinguish between clean and corrupted image regions (Levin, 2007; Chakrabarti et al., 2010; Schelten and Roth, 2014). Let us next consider different schools of thought to image restoration and their intrinsic challenges.

Deterministic Variational Approaches. Deterministic approaches rooted in the calculus of variations model the observed and unknown images as spatially continuous functions $u, f : \Omega \subset \mathbb{R}^2 \to \mathbb{R}$, respectively. The task of image restoration is then formulated as minimizing an energy functional of the form

$$\mathcal{E}(f;u) = \int_\Omega (u - Kf)^2 + \lambda \int_\Omega \varphi(\nabla f), \tag{1.4}$$

where K is a linear convolution operator and φ denotes a smoothing function (Aubert and Kornprobst, 2006). The minimum is searched for in a predefined (*e.g.*, Sobolev) space. Solutions necessarily fulfill the associated Euler-Lagrange partial differential equation (PDE), but solving it is generally intractable analytically. Hence stable numerical discretizations must be found, which is often challenging (Pock et al., 2007). In the case that φ is the well-understood total variation penalty $|\nabla f|$, and if the functional of Equation (1.4) is discretized directly by finite differences (FDs) or finite elements (FEs), a profound solution from convex analysis is Chambolle's minimization method (Chambolle, 2004; Zach et al., 2007; Aubert et al., 2009).

Bayesian Methods. One way to address the inherent uncertainty of inverse problems in computer vision is to use probabilities, *i.e.*, the unknowns are regarded as random variables following some distribution. In general terms, the object of interest is the posterior probability of the unknowns given the measured variables, *e.g.*, $p(\mathbf{x}, \mathbf{k}|\mathbf{y})$ for the case of blind deconvolution. The posterior may be recast using Bayes' theorem as a product of likelihood and prior terms, that is,

$$p(\mathbf{x}, \mathbf{k}|\mathbf{y}) \propto p(\mathbf{y}|\mathbf{x}, \mathbf{k})p(\mathbf{x}), \tag{1.5}$$

assuming a non-informative prior on the kernel, *i.e.*, $p(\mathbf{k}) \propto 1$, and statistical independence of the unknown image and blur. The likelihood $p(\mathbf{y}|\mathbf{x}, \mathbf{k})$ is derived from a forward model, which is relaxed to observe random noise deviations typically modeled as Gaussian (Equation (1.1)). Note that we have already discussed the difficulty of accurately identifying the image formation process in an earlier paragraph. On the other hand, the search for good image priors $p(\mathbf{x})$ is also severely challenging. To begin with, the derivative histograms of natural images have a higher kurtosis than the normal distribution (Huang and

Mumford, 1999; Huang, 2000), a property which manifests itself in a more tightly peaked and heavy-tailed curve as compared to the standard Gaussian bell shape. Thus, computationally convenient, spatially uniform Gaussian priors on image filter responses, which also underpin the Wiener filter variant of Equation (1.3), are too crude to yield high-grade results in practice. Markov random fields (MRFs) (Besag, 1974; Geman and Geman, 1984) are perhaps the most widespread type of prior in use today. These undirected graphical models impose local constraints on lattice-like data and are hence ideally suited for image processing. MRFs are typically determined by potential functions exerting unary influence or acting on the responses of image filters covering pairwise or higher-order (*e.g.*, 15×15) neighborhoods (cliques). In the rudimentary case study leading to Equation (1.3), the filters are typically hand-picked as simple, *e.g.*, 2-tap derivatives with squared norm potentials. By contrast, modern research papers *learn the filters and / or potentials from natural image data* (Zhu and Mumford, 1997; Roth and Black, 2005; Weiss and Freeman, 2007; Gao and Roth, 2012), which leads to vast improvements in quantitative experiments and a better understanding of images and their inherent statistical properties[4]. The complexity of prior learning may be attributed to the cumbersome, high dimensionality of the underlying image space, which makes it difficult to handle computationally (Weiss and Freeman, 2007; Keogh and Mueen, 2010).

With regard to the question of inference, given a loss (*i.e.*, cost) function $\mathcal{L}$, it is consistent with Bayesian decision theory to determine the optimal estimate $\widehat{\mathbf{x}}$ by minimizing the expected loss under the posterior distribution, *e.g.*,

$$\widehat{\mathbf{x}} = \operatorname*{argmin}_{\widetilde{\mathbf{x}}} \int \mathcal{L}(\widetilde{\mathbf{x}}, \mathbf{x}) p(\mathbf{x}|\mathbf{k}, \mathbf{y}) \, d\mathbf{x}, \tag{1.6}$$

in case the blur is also known. For the squared loss $\mathcal{L}(\widetilde{\mathbf{x}}, \mathbf{x}) = \|\widetilde{\mathbf{x}} - \mathbf{x}\|^2$, which also lies at the heart of the well-known peak signal-to-noise ratio (PSNR), the Bayes optimal estimate is the posterior mean. This is also known as the minimum mean squared error (MMSE) estimate. For structural similarity index (SSIM), a more advanced image metric, no similar relation is known. It quickly becomes evident that approximate algorithms to solve Equation (1.6) are required for all but the simplest posterior densities and loss functions. One option is to use Markov chain Monte Carlo (MCMC) techniques such as Gibbs sampling (Geman and Geman, 1984; Jensen et al., 1995) and essentially perform integration by averaging over samples (Schmidt et al., 2010, 2011; Sollweck, 2012). This approach holds the promise of high accuracy, but the required computing time is often prohibitive, *i.e.*, it is challenging to find efficient sampling schemes. A different approach is termed variational Bayesian inference (Jordan et al., 1999; Winn

4 These image models are also called *generative*.

and Bishop, 2005), and consists of fitting a simpler, surrogate density to the original one (Miskin and MacKay, 2000; Schelten and Roth, 2012; Wipf and Zhang, 2014). In case the approximation factorizes in every variable, we speak of the mean field procedure. Variational methods are often more efficient than sampling, but the challenge lies in maintaining control over the error arising by design from the less expressive functional form of the approximate distribution.

Discriminative Methods. As noted above, it is highly complex to conceive viable model forms and accompanying learning schemes for natural image priors. Although research into generative image modeling has progressed far, computational efficiency often remains an issue, while samples drawn from prior distributions still differ substantially from the appearance one might expect of natural scenes (*e.g.*, Levi, 2009). In contrast to Bayesian methods, the discriminative approach obviates the need for image priors by striving to directly determine the posterior probabilities, *e.g.*, $p(\mathbf{x}|\mathbf{k}, \mathbf{y})$, for the case of non-blind deconvolution. The underlying principle is formulated by Vapnik (1998), *"If you possess a restricted amount of information for solving some problem, try to solve the problem directly and never solve a more general problem as an intermediate step"*. Of course, modeling the posterior is also a very difficult undertaking. A popular approach is to use conditional random fields (CRFs). These are MRFs whose potential functions are determined from the input data, *e.g.*, the corrupted image and blur kernel. However, finding an effective mapping from image data to potentials is not easy. One successful approach is to use decision trees (Nowozin et al., 2011). To guarantee efficient learning and inference, Jancsary et al. (2012a,b) further propose to restrict the CRF class by regressing the parameters of *Gaussian* random fields from the input image. The resulting model is known as Regression Tree Fields (RTFs). With regard to training, discriminative approaches are often powerful enough to completely memorize data. If proper precautions are not taken, this can lead to overfitting on the training set, *i.e.*, poor generalization to new data. Learning may be particularly problematic if the sample size of training data is too small to adequately represent the variability encountered in real-world applications. To counteract overfitting, the model parameters must be calibrated, *e.g.*, the decision tree parameters used for the RTF must be chosen with care using measures such as cross-validation. On the other hand, aforementioned adaptability to the training data clearly offers leverage. In image deconvolution, it can lead to effective mitigation of artifacts caused by kernel estimation defects (Schmidt et al., 2013). Correspondingly, an alternating cascade of kernel estimation and image prediction is more than the sum of its parts when trained discriminatively (Schelten et al., 2015).

1.4 CONTRIBUTIONS

The goal of this thesis is to advance the state of the art in low-level vision[5], particularly in the field of image restoration. Broadly speaking, we make contributions to the areas of modeling and inference in image restoration, while a recurring theme throughout this dissertation is the open problem of image deconvolution. In more detail, the contributions of this thesis may be summarized as follows.

- With regard to model foundations, we contribute towards reconciling two seemingly conflicting schools of thought, deterministic variational models and probabilistic MRFs by establishing a connection between them. The MRF structure arises by restricting the variational functional to linear combinations of finite elements with local support and explicitly integrating over the domain variable. It is important to note that similar relations have been shown by Szeliski (1990a). In contrast to previous work, however, we focus on variational models having modern, non-quadratic regularizers, which generally exhibit higher accuracy in low-level vision applications due to their resistance towards outliers.

- We give a detailed study of a Gibbs sampling-based, Bayesian approach to non-blind deblurring. The results confirm the advantageous effect of MMSE estimation for a generatively learned, high-order prior (Schmidt et al., 2010). Please note that the integrated noise estimation is *not our contribution*, but of Schmidt (2016).

- We investigate fully factorized mean field inference for continuous-valued, high-order MRFs, and find that for a class of MRFs with Gaussian scale mixture (GSM) potentials, the updates scale quadratically in the clique size. We contribute an extensive experimental validation on a diverse array of applications spanning denoising, noise estimation, non-blind deblurring, and layer separation.

- We present a new approach for blind removal of localized blur, which is caused by fast object motion in the scene or camera defocus. The model employs a pixel-wise hidden variable to switch between blurs to explain different areas across the image plane. Another feature is non-parametric kernel estimation, which leads to improved generalization and is different from most previous methods (*e.g.*, Levin, 2007; Chakrabarti et al., 2010).

5 Low-level computer vision is an umbrella term comprising processing tasks such as image restoration, edge detection, segmentation, and optical flow.

- We introduce a new method for blind deblurring of camera shake based on interleaving discriminatively trained, sharp image predictions with kernel re-estimation steps. The discriminative predictions are carried out using the RTF model of Jancsary et al. (2012a,b), which is partly trained using a new set of realistic blur kernels from human hand motion. The kernel data was recorded at Microsoft Research Cambridge and *is not a contribution* of this dissertation.

1.5 THESIS OVERVIEW

The content on the upcoming pages of this thesis is structured as follows.

- Chapter 2 gives an overview of the related work relevant to this thesis. The literature covered here includes foundations of low-level vision, approximate probabilistic inference methods, as well as specific foundations and specialized techniques for image deblurring.

- Chapter 3 is concerned with model foundations of image restoration. In particular, we draw a link between MRFs and variational approaches. This chapter is based on the paper "Connecting non-quadratic variational models and MRFs" (Schelten and Roth, 2011a) presented at CVPR. Further material was previously published in a tech report of the same title (Schelten and Roth, 2011b).

- Chapter 4 demonstrates non-blind deblurring via Gibbs sampling-based Bayesian inference. This chapter corresponds to the paper "Bayesian deblurring with integrated noise estimation" (Schmidt et al., 2011) presented at CVPR. As noted in Section 1.4, integrated noise estimation is not our contribution, but of Schmidt (2016).

- Chapter 5 contains a study of efficient Bayesian inference for MRFs with high-order cliques based on the mean field approximation. Applications are shown in denoising, noise estimation, non-blind deblurring, and layer separation. This research also appeared at DAGM as "Mean field for continuous high-order MRFs" (Schelten and Roth, 2012).

- Chapter 6 introduces a new model for localized blur removal. This work was published at ICPR under the title "Localized image blur removal through non-parametric kernel estimation" (Schelten and Roth, 2014).

- Chapter 7 puts forward a discriminative method for blind deblurring. This material corresponds to the paper "Interleaved

regression tree field cascades for blind image deconvolution" (Schelten et al., 2015) presented at WACV.

- Chapter 8 provides a summary as well as an outlook on the further development of the ideas presented in this dissertation.

2

RELATED WORK

CONTENTS

To reiterate from Chapter 1, *image restoration* refers to the task of recovering a clean image given an input suffering from data faults such as flawed pixels, sensor noise, and blur of various sources. The nature and limitations of the imaging process dictate that such defects cannot be precluded entirely. However, there is a widespread need for accurate imaging across many areas of science and engineering. This has led to a large body of previous work in image restoration, which would be impossible to enumerate exhaustively in this dissertation. Instead, the goal of this chapter is to embed the thesis into an appropriate context of related publications. We begin by reviewing previous work on variational models in Section 2.1, placing a focus on their discretization. Section 2.2 gives an overview of pertinent literature on Markov random fields (MRFs), specifically concerning high-order generative models and Bayesian inference techniques. Finally, Section 2.3 covers related work in image deconvolution, including the more specialized areas of *localized* and *discriminative* deblurring. Please note that the topical organization of literature chosen below is nonexclusive in the sense that different approaches may share characteristics across categories.

2.1 VARIATIONAL MODELS

BACKGROUND. Variational models typically incorporate *(1)* a data fidelity term, and *(2)* an extra regularization term to mitigate ill-posedness. Using an additional regularization term is a general mathematical technique (Tikhonov and Arsenin, 1977) that is particularly suited to image restoration, which abounds with underdetermined problems. Early regularization terms on the image gradients tended to be quadratic, thus poorly representing the statistics of natural images and leading to unnaturally oversmoothed images. To address

this, Rudin et al. (1992) put forth the total variation (TV) model for image reconstruction, which combines a quadratic data term with the L^1-norm penalty on the image gradients. We also refer to this as the Rudin-Osher-Fatemi (ROF) model, and it remains popular today, since it preserves image edges, and due to its convexity, gives access to a rich toolbox of optimization methods. A variation of the ROF functional generalized to other modern penalties forms the basis of our argumentation in Chapter 3. Nevertheless, one negative effect of TV regularization is *staircasing*, which designates the phenomenon of smooth slopes in the true signal being reconstructed as piecewise constant. An effective countermeasure is to additionally regularize higher derivatives (Chambolle and Lions, 1997; Chan et al., 2000; Bredies et al., 2010; Chan et al., 2010).

SOLUTION METHODS. One approach to solving variational problems is to explicitly evaluate the variational functional for a general linear combination of basis functions and minimize the resulting expression w.r.t. the basis coefficients (Ritz, 1909; Bathe, 2014). The *Ritz method* is an integral part of our approach in Chapter 3. Another possibility is to derive and solve the associated Euler-Lagrange partial differential equation, which any minimizer must necessarily fulfill (Gelfand and Fomin, 2003; Evans, 2010). As a general remark, we note that while variational functionals are formulated for functions defined on the continuous domain, any solution approach on a digital computer must involve some form of discretization. In the following, we consider two discretization methods, finite differences (FDs) and finite elements (FEs). Both types of discretization often lead to large, sparse systems of linear equations, motivating specialized preconditioning schemes (Szeliski, 1990b, 2006; Pentland, 1994; Chan et al., 1999; Krishnan et al., 2013) to compute iterative solutions (Saad, 2003).

Finite Differences. The finite difference method (FDM) reduces continuous functions to a grid of values sampled at regularly spaced intervals in the domain, and the derivatives are expressed by finite difference approximations using the discrete grid (Grossmann et al., 2007). Solving the variational problem equates to searching for the optimal set of function values. Let us consider instructive examples of this technique. Chambolle (2004) discretizes the TV functional by finite differences and then introduces a minimization algorithm with proven convergence to the exact optimum: By use of a dual variable, minimizing the energy is shown to correspond to a nonlinear projection, which is in turn implemented by an efficient fixed point algorithm. This fast, duality-based procedure is embedded into a real-time optical flow method by Zach et al. (2007). To give another example for the usage of finite differences, Pock et al. (2007) apply algorithmic differentiation to the computer programs implementing dis-

cretized variational objectives. This yields higher-order derivatives, which permits the use of fast Newton-type optimization. Grabner et al. (2008) also utilize this approach for medical registration with graphics processing unit (GPU) acceleration.

Finite Elements. The finite element method (FEM) restricts the space of admissible functions to a linear basis of typically polynomial, locally defined functions termed finite elements (Courant, 1943; Bathe, 2014). Conceptually, the variational problem is broken down to a more tractable, finite-dimensional form, with the basis coefficients as the unknowns. Besides the standard polynomial elements, radial basis functions are a popular choice (Broomhead and Lowe, 1988; Buhmann, 2000; Morse et al., 2005; Gelas et al., 2007). Let us consider informative examples of this technique in computer vision. In multiresolution surface interpolation from noisy depth measurements with missing data, as may arise *e.g.*, from structured light or stereo procedures, the work of Terzopoulos (1983, 1984, 1986, 1988) illustrates a variational approach with *controlled-continuity* terms allowing for surface discontinuities, *i.e.*, the resulting FE discretization is piecewise smooth. For application in superresolution or denoising, Viola et al. (2012) acquire a resolution-independent, *i.e.*, continuously defined, latent image by fitting a regularized, piecewise linear FE element approximation to the input pixels. The fit is computed by minimizing an energy functional w.r.t. not only the FE parameters, but also the vertices and edges of the triangle mesh. This approach is further improved by MacCormick and Fitzgibbon (2013), who add curvature regularization to the latent image computation.

2.2 MARKOV RANDOM FIELDS

In contrast to variational models operating on the continuous domain, MRFs model spatially discrete random variables. An MRF fulfills the property that any variable is conditionally independent of the others given its set of neighboring variables as specified by an undirected graph (Li, 2009; Barber, 2012). The graph describes the independence structure of a distribution. We note that the variables may have both discrete or continuous values. One virtue of MRFs is the *"knock-on effect"* (Blake et al., 2011), which denotes that interactions between variables need only be formulated over short distances, but information is transmitted over longer ranges. Another benefit is the number of available inference methods, among which belief propagation (Pearl, 1982, 1988; Yedidia et al., 2003; Felzenszwalb and Huttenlocher, 2006) and graph cuts (Boykov et al., 2001; Kolmogorov and Zabih, 2004; Rother et al., 2004) feature prominently. The ensuing lines will place a particular emphasis on certain instances of Markov chain Monte Carlo (MCMC) and variational inference.

BACKGROUND. The Ising model of ferromagnetism (Lenz, 1920; Ising, 1925) is perhaps the first example of a MRF. Further, Besag (1974) studies MRFs as conditional probability models consisting of spatially distributed variables on a lattice with applications in plant ecology. This work establishes important foundations with contributions including a proof of the Hammersley-Clifford theorem (Hammersley and Clifford, 1971; Clifford, 1990), the introduction of Gaussian MRFs (Rue and Held, 2005), and early insights into approximate inference techniques. Due to the local similarities between pixel values as well as the spatial grid structure of image data, MRFs are a natural choice for modeling images. This was recognized early on by Geman and Geman (1984), who introduce a MRF image prior and the *Gibbs sampling* algorithm, which belongs to the class of MCMC procedures to sample from complex distributions (Gelfand and Smith, 1990; Neal, 1993). It is a viable approach if samples can be drawn efficiently from the conditionals of a probability. In its basic form, the algorithm updates each variable by sampling conditioned on the others. In the limit, the samples will stem from the joint distribution. The algorithm is often run for several burn-in steps and samples are only extracted every n-th cycle to reduce dependencies. One approach to increase efficiency consists of sampling groups of variables simultaneously. This is termed block Gibbs sampling (Jensen et al., 1995; Jensen and Kong, 1999). Let us continue with a discussion of MRFs as generative models for natural images in the next paragraph.

LEARNED HIGH-ORDER MRF PRIORS. Handpicked constraints on the responses of two-tap derivative filters give rise to simple MRF image priors with pairwise cliques. However, further ranging connections clearly hold the potential for richer models and are in fact used to capture higher frequency content such as image texture. The immediate notion of augmenting by higher than first-order derivatives naturally increases expressiveness, yet still entails hand-selection of model components. An effective measure to limit such manual input is to learn high-order MRF priors from image data, which is often plentiful. Let us here focus on previous work of this type. For application in texture modeling and synthesis, Zhu and Mumford (1997); Zhu et al. (1997, 1998) propose the Filters, Random Fields and Maximum Entropy (FRAME) model, which is characterized by discrete-valued potentials learned from observed filter responses through the maximum entropy principle (Jaynes, 1957). Note that the filters are selected automatically during training from a preconfigured bank of image filters containing, *e.g.*, Gabor filters (Gabor, 1946; Daugman, 1985). Beyond texture images, the general-purpose Product of Experts (POE) framework of Hinton (1999, 2002) captures natural image patches by a product of distributions *(experts)* on the scalar responses of filters (Welling et al., 2003; Teh et al., 2003). POE potentials are inte-

grated into a MRF prior on entire images by the Field of Experts (FOE) model (Roth and Black, 2005, 2009), which admits the training of both potentials and filters from natural image data. Weiss and Freeman (2007) advance likelihood bounds for FOEs with Gaussian scale mixture (GSM) potentials, a probability class comprising the previously used Student-t distribution. Using an auxiliary-variable block Gibbs sampler (Levi, 2009), it is possible to train FOE image priors with GSM potentials and achieve high-grade restoration performance in peak signal-to-noise ratio (PSNR) by averaging over samples to approximate the posterior mean (Schmidt et al., 2010). This equates to computing the Bayesian minimum mean squared error (MMSE) estimate, and we extend this approach to non-blind deblurring in Chapter 4. For the same model form, Gao and Roth (2012) obtain enhanced performance in sample statistics and image denoising by introducing refinements of the learning procedure, such as circular-boundary, *toroidal* sampling, which induces an accelerated exploration of the sample space.

VARIATIONAL INFERENCE. For many MRFs in computer vision, determining optima or expected values is intractable due to the complexity of the graph structure and the high dimensionality of the underlying image space. This motivates the use of approximate inference methods. *E.g.*, as discussed in the previous paragraph, expectations may be approximated by averaging over samples from a MCMC method. However, determining when the Markov chain has reached the desired distribution is a challenging problem, and required runtime is often prohibitive. By contrast, let us here review deterministic, variational approaches. These follow the strategy of approximating the intractable model by a simpler proxy distribution, which is then used for further reasoning (Neal and Hinton, 1999; Jordan et al., 1999; Jaakkola, 2001; Wainwright and Jordan, 2008). The approximation is calculated by minimizing the Kullback-Leibler (KL) divergence between the proxy and the true probability, a requirement that also admits message passing schemes (Winn and Bishop, 2005; Minka, 2005). The standard approach is to assume a surrogate distribution consisting of separate factors for each random variable. This method is known as *mean field* and originates from statistical physics (Parisi, 1998). In the remainder of this paragraph, we will list various instances of the technique in computer vision. For low-level applications such as surface reconstruction from sparse and irregular depth data, early work of Geiger and Girosi (1991) employs variational inference complemented with the additional annealing technique of gradually lowering the system temperature to find improved optima (Kirkpatrick et al., 1983). Iteratively decreasing the noise estimate is also an effective measure for kernel estimation in blind deblurring, where variational methods are used to increase robustness (Miskin

and MacKay, 2000; Fergus et al., 2006; Levin et al., 2011; Babacan et al., 2012; Wipf and Zhang, 2014). In Chapter 6, we broaden this paradigm to the removal of localized image blur. Moreover, Weinman et al. (2008); Pal et al. (2012) show how to use efficient, *sparse* variational updates to learn conditional random field (CRF) models for stereo. To further give an example of complex graph structure, previous work advanced by Krähenbühl and Koltun (2011) pertains to CRFs with Gaussian feature space potentials acting on all possible pairs of pixels in the image. Approximate variational inference is performed efficiently by Gaussian filtering. Vineet et al. (2014) extend this inference type to high-order models and product label spaces. We note that both above approaches are presented in the context of discrete-valued problems such as image segmentation and labeling, whereas in Chapter 5, we consider mean field as a fast method for inference in continuous valued, high-order MRFs using image restoration as a testbed.

CONNECTIONS TO VARIATIONAL MODELS. Having listed significant previous work on MRFs, let us now turn to linking them with the variational models covered in Section 2.1. It is not hard to intuit similarities between these approaches, yet the question is how to consolidate such connections. Thereby, the difficulty lies in overcoming the gap between rather unalike concepts such as the use of spatially discrete random variables on the one hand, and continuous-domain functions on the other. There is little previous research stating unequivocal model links. A definite exception is the work of Szeliski (1990a), who for low-level vision applications such as structure from motion, considers variational energy functionals defined with quadratic regularizers (the *membrane* and *thin-plate* functionals, see Terzopoulos, 1986, 1988) and deduces a Gaussian MRF formulation by virtue of discretizing with FEs. The associated covariance matrix allows to additionally quantify the uncertainty of the computed estimates, which is, *e.g.*, useful information for camera motion estimation (Szeliski, 1988, 1990a). By contrast, Chapter 3 of this thesis goes beyond the above work to establish model connections for more advanced, non-quadratic penalty functions, which preserve edges and lead to considerably more accurate results in image restoration, but are generally more challenging w.r.t. optimization and integration as required by a Ritz approach.

2.3 IMAGE DEBLURRING

While the previous two sections covered aspects of variational models and MRFs, we now take a closer look at the particular task of image deblurring. The problem is generally formulated as extracting an unknown signal from a given convolution with another, where the

latter is also unknown in the more difficult, and arguably more realistic, *blind* scenario. Such *deconvolution* problems are not restricted to camera sensors, but occur in many areas of signal analysis. Examples include the enhancement of resolution in seismic data (Wiggins, 1978; Mendel, 1983), and the removal of reverberations in acoustics (Bell and Sejnowski, 1995; Naylor and Gaubitch, 2010). Specifically, the problem statement for image deblurring is to find a sharp explanation of an image degraded by a potentially unknown blur and further corrupted by noise (Kundur and Hatzinakos, 1996a,b; Campisi and Egiazarian, 2007; Chaudhuri et al., 2014; Wang and Tao, 2014). In the following, we discuss selected solution techniques to convey the most important concepts in preparation for the upcoming chapters.

2.3.1 *Fundamentals*

Publications on image deblurring can be sorted into two problem categories, namely *non-blind* and *blind*. The ensuing text covers related work for both, with the objective of laying the groundwork for a discussion of more specialized research on localized and discriminative methods in Sections 2.3.2 and 2.3.3.

NON-BLIND DEBLURRING. The problem of restoring a sharp image from a blurred one *given the blur kernel* is known as non-blind deblurring. Using a Bayesian formulation with Poisson shot noise assumption, Richardson (1972) and Lucy (1974) iteratively calculate a maximum likelihood estimate. Although fast and widely used, the Richardson-Lucy algorithm frequently amplifies noise and leads to unacceptable ringing artifacts in the deblurring result. By contrast, Levin et al. (2007) obtain much crisper results using a natural (peaky and heavy tailed) image prior. The authors further elucidate how non-blind deblurring cannot generally equate to straightforward kernel inversion due to possible zeros in the kernel spectrum and blow-up of additive noise contributions at kernel frequencies of small magnitude. Nevertheless, it must be mentioned that the shape of natural image priors leads to more involved solution procedures. One option to cope with this challenge is the approach of Krishnan and Fergus (2009), who use a fast type of half-quadratic minimization (Geman and Reynolds, 1992; Geman and Yang, 1995), which iterates between solving a circulant system of linear equations by fast Fourier transform (FFT), and a pixel-wise, one-dimensional minimization problem by use of look-up tables or analytically in special cases. Regarding the trade off between runtime and precision, our contribution of Chapter 4 is at the opposite end of the spectrum. We show that in combination with a generative prior, excellent results in non-blind deblurring are achieved by Gibbs sampling-based Bayesian estimation (Schmidt et al., 2011). While our results are based on FOE priors for the en-

tire image, Zoran and Weiss (2011, 2012) learn a Gaussian mixture model (GMM) for *patches* of natural images, which, when integrated into a general framework combining a data formation term with the Expected Patch Log Likelihood (EPLL), yields state of the art non-blind deblurring results. Note that we will review recent, *discriminative* methods for non-blind deblurring in Section 2.3.3.

BLIND DEBLURRING. Reversing image blur without knowledge of the blur kernel is referred to as blind deblurring. Early work of Miskin and MacKay (2000) treats the task of separating a weighted sum of blurred cartoon images, which is achieved by a variational approximation to the true posterior. However, this simple class of images is modeled as having independent, identically distributed pixels, an assumption which is limited in scope. In the course of adapting the latter work to color photographs, Fergus et al. (2006) determine an influential algorithmic paradigm, with the most widely adopted components being gradient domain, coarse-to-fine kernel estimation, accurate image prior modeling, and the observation of variable uncertainties. Going further, the theoretical underpinning for many blind methods, including the two already mentioned in this paragraph, is provided by Levin et al. (2009, 2011). The authors investigate the shortcomings of optimizing simultaneously w.r.t. both the hidden image and blur kernel and use estimation theory to advocate estimating the kernel after marginalizing over the unknown image variables[1]. Typically, the marginalization is carried out by variational inference. Wipf and Zhang (2014) clarify how this leads to a particularly robust maximum a-posteriori (MAP) problem. With regard to a TV image prior, Perrone and Favaro (2014, 2016) show that the key to success lies in detaching the application of non-negativity and normalization constraints on the blur kernel during optimization. Further, we note that the difficulty of blind deblurring lies partly rooted in the fact many image priors favor blurred images over sharp ones. Krishnan et al. (2011) address this with an improved regularization term, which assigns higher cost to both blurry and oversharpened images. The proposed gradient measure is the L^1-norm divided by the L^2-norm. This expression has the decisive property that blur decreases the denominator at a faster pace than the numerator, thus increasing the cost and guiding the optimization away from the blurred image.

Edge Maps. Several previous publications limit the kernel computation to image areas containing critical gradient information, thus making gains in runtime and robustness. *E.g.*, for lesser blur distortion, Joshi et al. (2008) reconstruct step edges at selected locations in

1 Moreover, the publication includes a widely used benchmark data set of test images for the quantitative evaluation of camera shake removal. Further available benchmarks are from Köhler et al. (2012); Sun et al. (2013); Mai and Liu (2015); Lai et al. (2016).

the image and use the predicted pixels to estimate the kernel. In a similar vein, Cho et al. (2011) recover small-scale kernels by estimating their Radon transform from the deformation of step edges. Coping with larger blur, Cho and Lee (2009) compute the kernel from relevant edge locations obtained from the current image estimate by application of bilateral filtering (Tomasi and Manduchi, 1998), shock filtering (Osher and Rudin, 1990), and thresholding of the gradient magnitude. This permits to estimate the hidden variables using simple Gaussian priors, which leads to fast processing in the frequency domain. Xu and Jia (2010) develop a related approach based on masking out image areas with finer structure for improved blur prediction. The expensive sparsity requirement on the kernel is imposed as a separate step following the multiscale routine. We use this two-phase algorithm to initialize the deconvolution procedure of Chapter 7. Later work of Xu et al. (2013) obviates the shock filtering and instead pursues a more principled L^0 optimization to obtain salient gradients for kernel estimation. Further, Sun et al. (2013) estimate the blur and latent image by reconstructing edge, junction, and corner structures using a patch prior of image primitives.

Non-Uniform Blur. Although the uniform convolution model utilized by most deblurring methods has computational advantages, in practice, camera motion causes non-uniform blur. *E.g.*, for in-plane rotational motion of the camera about the optical axis, the further the image region lies from the center point, the more severe the blur degradation becomes. This has been addressed by several authors (Whyte et al., 2010; Gupta et al., 2010; Tai et al., 2011), who model the image formation as a weighted sum of sharp image homographies. Thereby, each weight encapsulates the time the camera spends in a particular pose stemming from a discretized space of possible motions. Note that the weight set generalizes the conventional concept of the convolutional blur kernel. However, the calculation of image homographies is computationally expensive. One remedy is the Efficient Filter Flow (EFF) framework (Hirsch et al., 2010), which recreates smoothly varying blur by overlapping uniformly convolved regions and leads to considerable speed-ups in practice (Harmeling et al., 2010; Hirsch et al., 2011; Whyte et al., 2011; Xu et al., 2013). Finally, we note that for small camera motion, it suffices to estimate locally linear kernels (Kim and Lee, 2014; Sun et al., 2015).

Inertial Sensors. In contrast to purely image-based approaches, Joshi et al. (2010) leverage information from motion sensors affixed to the camera. In particular, gyroscopes and accelerometers yield measurements which, when integrated over time, allow to estimate the camera's relative rotational and translational motion. With these, it is feasible to derive image homographies expressing spatially variant

blur. The authors address the inherent problems of sensor drift and scene depth estimation by computing the motion which is optimal under a joint score of observation model and latent image prior. Similarly, Horstmeyer (2010) utilizes inertial measurement devices fastened to the camera, while Šindelář and Šroubek (2013); Šindelář et al. (2014a,b); Hu et al. (2016) rely on built-in smartphone motion sensors.

2.3.2 *Localized Deblurring*

Beyond camera shake degradation, fast moving objects in the scene may also be rendered blurry, and defocus can arise in case of sufficiently large depth variations in the scene. We refer to this type of image corruption as *localized blur*, since the crucial difference to camera shake is that only restricted subsets of pixels are affected. Because of the spatially limited region of interest (ROI), camera motion deblurring algorithms fail on localized blur, motivating the need for specialized solutions.

USER ASSISTANCE. One solution approach is to query the user for brush strokes marking the blurred object and background. This permits to compute the object alpha matte (Levin et al., 2006, 2008a), which often tapers heavily toward the borders. In fact, Jia (2007) shows that the transparency at the boundary of motion blurred objects contains sufficient information to estimate the kernel and deblur the object. Besides brush strokes, Shan et al. (2007) also require user supplied points to identify and remove the blur of a rigid object undergoing rotational motion. To address both fore- and background blur, Dai and Wu (2009) propose a refined two-layer image recovery model which further reconstructs the alpha matte of the sharp object, albeit given the user initialization.

HARDWARE MODIFICATIONS. Conventional imaging technology can be modified to better cope with localized blur. To begin, Raskar et al. (2006) show that fluttering the camera shutter during exposure leads to a more benign blur of moving objects in the scene, in the sense that high frequencies can be reconstructed by deconvolution. Levin et al. (2007) overlay the traditional lens with an occlusion pattern. This generates defocus blur having a distinctive set of zeros in the frequency domain for different scene depths. It follows that an all-focus image can be recovered by non-blind deblurring with different scalings of the aperture filter. For objects moving horizontally at different velocities, Levin et al. (2008b) demonstrate that camera motion during exposure produces a uniform, high frequency conserving image blur. There is accordingly no need to segment the image into static background and moving objects for non-blind deblurring. Moreover, Martinello and Favaro (2011) restore a sharp image from

simultaneous motion and defocus blur using an aperture fragmentation that preserves richer frequency content than a standard lens.

AUTOMATIC IMAGE BASED METHODS. Beyond user interaction and hardware measures, there is also previous work relying solely on image input. One research direction goes into video data, which adds a temporal dimension to exploit. *E.g.*, for consecutive frames of a video with a moving object, Bar et al. (2007) propose a modified Mumford-Shah formulation to simultaneously segment the moving object, estimate its speed and remove the blur. Rather than video data, we will here focus on single image input.

Candidate Kernels. To facilitate the task of localized deblurring, one can approximate motion blur by box filters, and camera defocus by Gaussian filters. The optimal blur is then searched for among a finite set of options, *e.g.*, box filters from a matrix of speeds and orientations, and/or Gaussian filters spanning a range of strengths. One way to find the responsible blur is image statistics. Levin (2007) matches per-filter derivative histograms to determine which kernel caused the localized blur. Chakrabarti et al. (2010) compute the most likely candidate blur under a sparse distribution of image gradients. The blurred object is segmented with a variant of *GrabCut* (Rother et al., 2004), *i.e.*, by graph cut minimization of an energy composed of the likelihood of a pixel being blurred or sharp, a GMM to describe the fore- and background color, and a simple Potts prior to enforce spatial coherence. As opposed to binary blur/no-blur segmentations, another line of work employs learned models to compute local blur likelihoods, which then serve as unary terms of a multilabel energy minimization problem. In particular, Couzinié-Devy et al. (2013) use dictionary and Gabor filters to extract feature vectors for every pixel, which are then provided as input to a learned, logistic regression model expressing the probability of each blur. As a general remark, we note that the search space of candidate kernel methods can be extended by rotating or scaling the image to estimate blurs outside of the specified set (Sun et al., 2015).

Freely Varying Kernels. Instead of positing the appearance of localized blur in advance, it is also feasible to calculate kernels with freely varying elements. In that regard, Kim et al. (2013) advance a convex optimization algorithm which estimates object motion and camera shake blurs, while simultaneously segmenting the image into the affected regions. In a different approach, Shi et al. (2014) develop blur detection features to identify the corrupted area, whereupon the blur can be removed locally with a uniform method; the proposed features are based on image derivative statistics, average power spectrum, and learned filters. In Chapter 6, we contribute a marginal like-

lihood, latent variable framework that is general enough to cope with both motion and defocus blur (Schelten and Roth, 2014).

2.3.3 *Discriminative Deblurring*

The vast majority of the previously referenced methods adhere to the Bayesian paradigm and reformulate the posterior using image likelihood and prior models. In recent years, however, promising research has been made into discriminative procedures for deblurring. These approaches directly model the posterior, and we will discuss non-blind and blind instances in this subsection.

NON-BLIND METHODS. Schuler et al. (2013) carry out a regularized inversion of the blur kernel (Wiener, 1964) and remove the resulting, heavy image corruption by patch-wise application of a multilayer perceptron learned on a large database of images. The drawback is that the neural network needs to be trained specifically for each input kernel, such that in practice, the method is only feasible in controlled scenarios with invariant blur. A more general approach, which also allows for unseen kernels, is based on the Regression Tree Field (RTF) model (Jancsary et al., 2012a,b). The latter defines a class of Gaussian CRFs having potentials regressed locally from the input image by decision trees. Training is conducted in a loss-based fashion, allowing for different image quality metrics. Schmidt et al. (2013, 2016) subsume the standard deblurring likelihood into a Gaussian CRF with unknown regularization parameters to be regressed by the RTF framework. Sharp image prediction proceeds iteratively over a cascade of RTFs, each receiving the previous result for refined feature extraction. Schmidt and Roth (2014) also rely on a cascade architecture, but propose the faster *shrinkage fields* model for prediction, which is derived from a variant of half-quadratic optimization and has an overall complexity of $\mathcal{O}(n \log n)$ in the number of pixels. The model achieves high-grade results by loss-based learning of its parameters, including those governing the shrinkage functions, which are responsible for reducing distortion such as noise.

BLIND METHODS. To extract image features advantageous to kernel calculation, Schuler et al. (2016) propose a layered neural network. This feature extraction is combined with simple kernel and latent image estimation steps to form one level of a stacked architecture, which may be learned in its entirety using back-propagation. In a different approach to the alternating estimation of image and kernel, Zuo et al. (2015) train per-iteration parameters of hyper-Laplacian gradient priors by loss minimization. Besides increasing performance, this frees the user of tediously hand-tuning parameters. Uniquely, Mai and Liu (2015) investigate how to merge several kernels from differ-

ent blind methods into an enhanced estimate. The authors rely on the RTF framework to not only fuse kernel values coordinate-wise, but also account for dependencies between adjacent blur elements.

Interleaved Cascades. Chapter 7 of this thesis proposes *interleaved* cascades of RTFs (Schelten et al., 2015), whereby the hidden image is recovered by a sequence of RTF predictions interwoven by, and custom trained to, kernel estimation steps. Furthermore, Xiao et al. (2016) advance an interleaved cascade of shrinkage fields expressly designed with high-order filters to address the particular statistics of text images, which tend to contain high frequency content vastly different from natural images.

3

CONNECTING NON-QUADRATIC VARIATIONAL MODELS AND MRFS

CONTENTS

Spatially discrete Markov random fields (MRFs) and spatially continuous variational approaches are ubiquitous in low-level vision, including image restoration, segmentation, optical flow, and stereo. Even though both families of approaches are fairly similar on an intuitive level, they are frequently seen as being technically rather distinct since they operate on different domains. In this chapter, we explore their connections and develop a direct, rigorous link with a particular emphasis on first-order regularizers. By representing spatially-continuous functions as linear combinations of finite elements with local support and performing explicit integration of the variational objective, we derive MRF potentials that make the resulting MRF energy equivalent to the variational energy functional. In contrast to previous attempts, we provide an explicit connection for modern non-quadratic regularizers and also integrate the data term. The established connection opens certain classes of MRFs to spatially-continuous interpretations and variational formulations to a broad range of probabilistic learning and inference algorithms.

3.1 INTRODUCTION

Many vision problems, particularly in low-level vision require some form of regularization and are posed as energy minimization problems. Two basic approaches dominate the literature: Spatially continuous variational approaches on one hand (Rudin et al., 1992; Schnörr et al., 1996), and spatially discrete MRF approaches on the other hand

(Besag, 1974; Geman and Geman, 1984). Variational approaches originate in the calculus of variations[1] and assume that, while camera sensors may only yield spatially discrete measurements, the underlying quantity of interest, such as the intensity of incident light on the sensor, is ultimately spatially continuous. MRF approaches take a different route: They assume that in order to be realizable on a digital computer, a finite representation needs to be ultimately used. Hence, the desired output as well as any intermediate representation are taken to be spatially discrete. Beyond that, variational models of low-level vision are typically seen as deterministic, while MRFs admit a probabilistic view and regard the measurements and the output (at least in the probabilistic case) as inherently uncertain.

An important question is thus whether and how these approaches can be related. Superficially, both formulations are clearly similar. Yet, variational models and MRFs seem to operate on different levels: Variational approaches involve functions on the entire domain, whereas MRFs impose local constraints through cliques of pixels. The question of their relationship has been posed since the early days of regularization, but explicit links have been demonstrated only for a few cases, such as for quadratic regularizers (Szeliski, 1990a). Hence, their technical relationship has remained somewhat nebulous, and even nowadays most work in low-level vision is firmly rooted in one or the other paradigm. In this chapter, we show that this need not be the case.

In particular, we address the widely perceived chasm between modern first-order variational and MRF approaches. Starting from a standard variational formulation for image restoration (Rudin et al., 1992), we show that if we minimize the spatially continuous energy functional over a finite element (FE) space of a certain degree, an explicit connection to an equivalent spatially discrete MRF can be established without having to approximate the spatially continuous energy functional. The nodes in the MRF are still continuous valued, however. The degree of the underlying FE representation in the variational formulation is directly linked to the size of the cliques in the equivalent MRF. In contrast to common FE approaches to variational problems, we do not perform additional discretization steps, but instead carry out an *explicit*, and usually *analytical integration* of the domain variable. Compared to previous work (Szeliski, 1990a), our approach not only applies to quadratic, but to a wide range of modern first-order regularizers. This link also demonstrates that certain MRFs have an explicit spatially continuous interpretation, which further tightens the connection between statistical and variational approaches (Scharr et al., 2003). Moreover, this opens variational approaches to direct and explicit probabilistic interpretations and to standardized infer-

1 The calculus of variations can also be used for approximate inference in probabilistic models. In this chapter, however, we take "variational" to mean non-probabilistic computer vision approaches based on spatially continuous energy functionals.

ence techniques, such as graph cuts (Boykov et al., 2001), belief propagation (Yedidia et al., 2003), or more advanced variants (Wainwright et al., 2005; Kumar et al., 2009). As an example, we demonstrate sampling the corresponding MRF prior and posterior (see Figure 3.3).

Beyond the main goal of establishing such a connection, our approach also compares favorably to the common finite difference (FD) method: Conceptually, our approach does not require to approximate the original spatially continuous energy functional; rather the functional is directly evaluated and minimized over a finite dimensional function space. Practically, our method improves the performance in an image restoration application.

3.2 BACKGROUND

VARIATIONAL APPROACHES. Variational approaches to low-level vision aim to minimize an energy functional that is comprised of a data fidelity term and a spatial term, *e.g.*:

$$\mathcal{E}(f;u) = \mathcal{E}_D(f;u) + \lambda\mathcal{E}_S(f) \tag{3.1}$$

$$= \int_\Omega (f-u)^2 \, d\mathbf{x} + \lambda \int_\Omega \varphi(\nabla f) \, d\mathbf{x}. \tag{3.2}$$

This example is a slight generalization of the standard Rudin-Osher-Fatemi (ROF) formulation for image restoration (Rudin et al., 1992). The data fidelity term $\mathcal{E}_D(f;u)$ ensures that the restored image $f : \Omega \to \mathbb{R}$, a spatially continuous function, is close to the noisy input $u : \Omega \to \mathbb{R}$, which is also assumed to be spatially continuous. Similar data fidelity terms can be used for other problems, such as optical flow (Papenberg et al., 2006). The spatial term $\mathcal{E}_S(f)$ regularizes the problem by encouraging spatial smoothness, and depending on the choice of $\varphi(\cdot)$ allows for discontinuities; λ controls the amount of regularization. Here we take image restoration as a representative example, but note that many variational formulations for optical flow, stereo, and other problems can be treated similarly.

There is a large variety of different approaches for solving variational problems. The traditionally most popular approach is to analytically derive the *Euler-Lagrange* partial differential equations and to subsequently discretize and solve them numerically (*e.g.*, Rudin et al., 1992; Papenberg et al., 2006). As noted by Pock et al. (2007), the derivation and solution of these equations can be tedious, and the resulting algorithm may not have an interpretation in terms of energy minimization. To address this, Pock et al. (2007) discretize the energy functional using FDs and propose minimization algorithms via algorithmic differentiation. Such finite difference methods (FDMs) have found widespread use (*e.g.*, Chambolle, 2004).

The third main category of solution methods, which we will rely on here, is based on finite element functions (Bathe, 2014). The finite

element method (FEM) uses local, usually piecewise polynomial representations for the sought after function. Like FD discretizations, FEMs approximately solve the original problem, but do so by restricting the space of functions over which the functional is minimized rather than approximating the functional itself, which can be intuitively viewed as a discretization of the function space. An advantage of FEMs is that different implementations can be compared in terms of the *original* spatially continuous energy. A variety of FE basis functions have been used, including standard finite element polynomials (Bathe, 2014), as well as B-splines (Ramani et al., 2007; Bernard et al., 2008) and radial basis functions (Morse et al., 2005; Gelas et al., 2007). While not as popular as FDMs, FEMs are also common in low-level vision. For example, Wang et al. (2006) use discontinuous FEs to perform image denoising while explicitly preserving edges. Vlasenko and Schnörr (2009) rely on a *mixed* FEM to derive a numerically stable discretization of a variational model for fluid flow estimation. In the context of surface reconstruction, Terzopoulos (1983) applies finite elements as discretization method to solve a corresponding variational principle. This work further investigates theoretical guarantees such as existence, uniqueness and convergence for both *conforming* and *nonconforming elements*, where the latter class is distinguished by relaxed differentiability requirements. Terzopoulos (1986) introduces spline functionals with weighting functions to allow for discontinuities, which enhances the precision of depth reconstruction. Furthermore, Terzopoulos (1988) demonstrates computationally efficient surface reconstruction from multiple scales with discontinuity preserving interpolation by use of a variational approach with ensuing finite element discretization.

In contrast to using the FE approach solely for the purpose of solving variational problems, the focus of this Chapter lies on linking modern variational models to MRFs.

MARKOV RANDOM FIELD APPROACHES. Most MRF approaches to low-level vision rely on a probabilistic formulation. Again using image restoration as an example (Geman and Geman, 1984), the posterior of the restored image[2] $\mathbf{o} \in \mathbb{R}^{N \times N}$ given the noisy input $\mathbf{i} \in \mathbb{R}^{N \times N}$ is formulated in terms of a likelihood and a prior:

$$p(\mathbf{o}|\mathbf{i}) \propto p(\mathbf{i}|\mathbf{o}) \cdot p(\mathbf{o}). \tag{3.3}$$

The likelihood $p(\mathbf{i}|\mathbf{o})$ is an observation model and intuitively takes the role of the data term from Equation (3.1). The prior $p(\mathbf{o})$ models the a-priori spatial regularity and intuitively corresponds to the spatial term from before. Note, however, that input $\mathbf{i}$ and output $\mathbf{o}$ are now

2 Note that images are to be interpreted as matrices or vectors throughout the thesis. As the correct interpretation is evident from the context, we have opted to keep the notation uncluttered.

spatially discrete and represented as arrays of $N \times N$ pixels; depending on the context, we also treat them as vectors. Please note that the arrays $\mathbf{i}$ and $\mathbf{o}$ denote the discrete analogs of the spatially continuous functions u and f; we use the different notation to explicitly illustrate that point.

The intuitive connection to variational models becomes even more apparent if we invoke the Gibbs measure notation

$$\begin{aligned} p(\mathbf{o}|\mathbf{i};T) &= \frac{1}{Z(\mathbf{i},T)} \exp\left\{-\frac{1}{T}E(\mathbf{o};\mathbf{i})\right\} \\ &\propto \exp\left\{-\frac{1}{T}\left(E_D(\mathbf{i};\mathbf{o}) + E_S(\mathbf{o})\right)\right\}, \end{aligned} \tag{3.4}$$

where $E(\mathbf{o};\mathbf{i})$ is the posterior MRF energy, and E_D and E_S are the respective MRF energies for the likelihood and the prior. The "temperature" T controls the shape of the distribution (for notational simplicity, we assume $T = 1$ in the following); the partition function $Z(\mathbf{i},T)$ ensures normalization. The goal of this chapter is to develop an explicit formal connection between the variational energy functional $\mathcal{E}(f;u)$ from Equation (3.1) and the posterior MRF energy $E(\mathbf{o};\mathbf{i})$.

The key property of MRFs, *e.g.* of the prior $p(\mathbf{o})$, is that a node (here, pixel) is conditionally independent of the rest (of the image) given its Markov blanket. By virtue of the Hammersley-Clifford theorem (Besag, 1974), the distributions factorize into products over the cliques $\mathcal{C}$ of the underlying graph

$$p(\mathbf{o}) = \frac{1}{Z}\prod_{c\in\mathcal{C}} \psi_c(\mathbf{o}_{(c)}) = \frac{1}{Z}\exp\left\{-\sum_{c\in\mathcal{C}} \rho_c(\mathbf{o}_{(c)})\right\}, \tag{3.5}$$

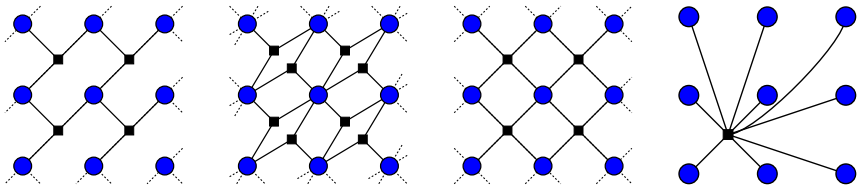
where $\mathbf{o}_{(c)}$ denotes all pixels belonging to clique c, $\psi_c(\cdot)$ are the factors, and $\rho_c(\cdot)$ are the corresponding potential functions. In the most frequent case of pairwise MRFs, the cliques consist of spatially neighboring pairs of pixels (Geman and Geman, 1984). In high-order MRFs (*e.g.*, Roth and Black, 2009), cliques consist of larger spatial neighborhoods of pixels.

A wide variety of different algorithms have been applied to MRF inference. Most often, the maximum a-posteriori (MAP) estimate

$$\mathbf{o}^* = \arg\max_{\mathbf{o}} p(\mathbf{o}|\mathbf{i}) = \arg\min_{\mathbf{o}} E(\mathbf{o};\mathbf{i}) \tag{3.6}$$

is computed. In the continuous-valued case[3], gradient methods are frequently used, which again bears similarities to the variational approach. In the discrete-valued case, belief propagation (Yedidia et al., 2003), graph cuts (Boykov et al., 2001), and more advanced variants (Wainwright et al., 2005; Kumar et al., 2009) have enjoyed enormous popularity in recent years. As an alternative to MAP, it is possible to compute the minimum mean squared error (MMSE) estimate, which

3 While MRFs are spatially discrete, they may well be continuous-valued.



(a) FDM (b) Linear FEM (c) Bilinear FEM (d) Biquadratic FEM

Figure 3.1: Factor graph structure of variational models. (a) Triple cliques of a standard FD discretization of the ROF model (Chambolle, 2004). (b) Triple cliques of linear FEs. (c) 2×2 cliques of bilinear FEs. (d) 3×3 cliques of biquadratic FEs (only one clique shown for clarity).

has been shown to outperform standard MAP estimation in the context of generatively trained models (Schmidt et al., 2010). Finally, we note that a statistical MRF formulation enables learning of model parameters (Roth and Black, 2009), which is paramount in complex models with many parameters.

3.3 CONNECTING VARIATIONAL MODELS AND MRFS

PREVIOUS WORK. The first connections between statistical models, such as MRFs, and variational problems have been developed early on. Szeliski (1990a) described a close relation between quadratic regularization and Gaussian MRFs using FEs. Starting from a first-order smoothness functional, the membrane model (Terzopoulos, 1986)

$$\mathcal{E}_S(f) = \frac{1}{2} \iint (f_{x_1}^2 + f_{x_2}^2) \, d\mathbf{x}, \tag{3.7}$$

Szeliski (1990a) uses linear FEs to derive the discretization

$$E_S(\mathbf{o}) = \frac{1}{2} \sum_{k,l} (o_{k+1,l} - o_{k,l})^2 + (o_{k,l+1} - o_{k,l})^2, \tag{3.8}$$

consisting only of forward differences. For image restoration, Szeliski (1990a), moreover, proposes the data term

$$\mathcal{E}_D(\mathbf{i}; f) = \frac{1}{2} \sum_m a_m \left(f(x_{1,m}, x_{2,m}) - i_m \right)^2, \tag{3.9}$$

which is discretized as

$$\begin{aligned} E_D(\mathbf{i}; \mathbf{o}) &= \frac{1}{2} \sum_{k,l} a_{k,l} \left(o_{k,l} - i_{k,l} \right)^2 \\ &= \tfrac{1}{2} (\mathbf{o} - \mathbf{i})^T A (\mathbf{o} - \mathbf{i}). \end{aligned} \tag{3.10}$$

The diagonal matrix $A = \operatorname{diag}\{a_{k,l} | k, l\}$ contains optional weights. Combining the discretized data and smoothness functionals results

in a quadratic form $E(\mathbf{o}) = \frac{1}{2}\mathbf{o}^T B\mathbf{o} - \mathbf{o}^T\mathbf{d} + e$. Invoking the Gibbs measure notation (see Equation (3.4)) yields the correspondence of the above functionals to a Gaussian MRF. These well-studied MRFs result in an elegant framework for computing estimates as well as their uncertainty from the posterior. Beyond this case of global quadratic regularizers, Kulkarni et al. (1994) furthermore relate MRFs and certain segmentation functionals.

While Szeliski (1990a) noted that non-quadratic regularizers can, in principle, be related to probability densities through Gibbs distributions, no explicit formulas for the potentials nor study of feasibility was provided. However, modern variational techniques tend to rely on non-quadratic penalties, which are robust toward outliers (Rudin et al., 1992; Aubert et al., 2009). Our key contribution is to revisit this relation and generalize it to the considerably more involved case of modern, non-quadratic regularizers, as well as to provide concise formulas for the equivalent MRF potentials.

FINITE DIFFERENCE APPROXIMATIONS. An approximate connection between variational models and MRFs, which quite surprisingly is rarely made, exists through finite difference discretizations. To be implemented on a computer, energy functionals have to be discretized somehow, which often happens using FDs (Chambolle, 2004; Pock et al., 2007). A standard FD discretization for the ROF model (Equation (3.1) with $\varphi(\mathbf{y}) = |\mathbf{y}|$) is to use forward differences. *E.g.*, Chambolle (2004) considers

$$E_{\text{ROF}}(\mathbf{o};\mathbf{i}) = \sum_{k,l}(o_{k,l} - i_{k,l})^2 + \lambda \sum_{k,l} \sqrt{(o_{k+1,l} - o_{k,l})^2 + (o_{k,l+1} - o_{k,l})^2}. \tag{3.11}$$

By defining appropriate prior potentials

$$\psi_c^S(\mathbf{o}_{(c)}) = \exp\{-\lambda\sqrt{(o_{k+1,l} - o_{k,l})^2 + (o_{k,l+1} - o_{k,l})^2}\} \tag{3.12}$$

with triple cliques $c = \{(k,l),(k+1,l),(k,l+1)\}$ as well as per-pixel likelihood potentials $\psi_{k,l}^D(i_{k,l};o_{k,l}) = \exp\{-(o_{k,l} - i_{k,l})^2\}$ an MRF whose energy is equal to Equation (3.11) is given as

$$p_{\text{ROF}}(\mathbf{o}|\mathbf{i}) \propto \prod_{k,l} \psi_{k,l}^D(i_{k,l};o_{k,l}) \cdot \prod_c \psi_c^S(\mathbf{o}_{(c)}). \tag{3.13}$$

Note that this MRF is not pairwise, since the prior potentials ψ_c^S involve 3 variables. The corresponding triple cliques are illustrated in Figure 3.1(a). One drawback of this FD approach is that different discretizations lead to different approximative functions, and thus cannot be compared in terms of the original variational objective. An implementation using finite elements with analytic integration, as is

pursued here, guarantees exact energy values in the sense that the *original* energy functional is evaluated instead of an approximation thereof.

FINITE ELEMENT (RITZ) APPROACH. Variational problems as in Equation (3.1) are typically formulated for infinite-dimensional function spaces $\mathcal{F}$, *e.g.*, the space of bounded variation (Aubert et al., 2009). To simplify the following treatment, we assume a generic variational model

$$\mathcal{E}(f;u) = \int_\Omega \mathcal{L}(f,u)\,\mathrm{d}\mathbf{x} \quad \to \quad \min_{f\in\mathcal{F}}, \tag{3.14}$$

where $\mathcal{L}$ is a real-valued expression depending on f, u, and derivatives thereof. To be implemented in practice, either the energy needs to be approximated in a finite-dimensional way (see Equation (3.11)), or the function space $\mathcal{F}$ over which the functional is minimized needs to be restricted. Such a smaller, restricted space of solutions is typically defined in terms of a finite set of basis functions $b_k(\mathbf{x})$; the functions are consequently determined by their basis parameters. In contrast to typical approaches (Ramani et al., 2007), we identify the basis parameters directly with the pixels, which means that the spatially discrete input and output images, $\mathbf{i}$ and $\mathbf{o}$, relate to the spatially continuous input and output functions, f and u, as

$$u(\mathbf{x};\mathbf{i}) = \sum_{k=1}^{K} i_k b_k(\mathbf{x}) \quad \text{and} \quad f(\mathbf{x};\mathbf{o}) = \sum_{k=1}^{K} o_k b_k(\mathbf{x}). \tag{3.15}$$

In this setting, the energy functional from Equation (3.14) is equivalent to a spatially discrete energy $E(\mathbf{o};\mathbf{i})$

$$\mathcal{E}(f;u) = \int_\Omega \mathcal{L}\big(f(\mathbf{x};\mathbf{o}), u(\mathbf{x};\mathbf{i})\big)\,\mathrm{d}\mathbf{x} = E(\mathbf{o};\mathbf{i}), \tag{3.16}$$

since the value of the functional only depends on the input and the output image. Such explicit integration is a classical approach and was described in an early work by Ritz (1909).

CONNECTION TO MRFS. To complete the connection to MRFs, we opt for basis representations with local support, such as FE representations (Bathe, 2014), which have been used in image restoration for a long time (Terzopoulos, 1988). The assumption of a basis representation with local control deserves some discussion: On one hand real optical systems pose limitations that make (piecewise) smooth functions appear very reasonable. On the other hand, variational problems with sub-linear regularizing terms[4] have been shown to be ill-posed (Aubert et al., 2009) in the space of functions with bounded

4 The regularizing term $\varphi(\cdot)$ is called sub-linear, if $\lim_{x\to+\infty} \frac{\varphi(x)}{x} = 0$.

variation, yet work well in practice when restricting the function space as done here.

We partition the domain $\Omega \subset \mathbb{R}^2$ into disjoint regions Ω_k (with $\Omega = \cup_k \Omega_k$ and $\Omega_i \cap \Omega_j = \emptyset, \forall i \neq j$), such that each region is overlapped with the same number of local basis functions (note that the regions Ω_k do not represent the support of the bases b_k). Then the values of the function on each region Ω_k are governed by those basis parameters that control the non-zero basis elements; we denote their indices as $c(k) \subset \{1, \dots, K\}$. Consequently, we have that

$$\mathcal{E}(f;u) = \int_\Omega \mathcal{L}\big(f(\mathbf{x};\mathbf{o}), u(\mathbf{x};\mathbf{i})\big)\, d\mathbf{x} \tag{3.17}$$

$$= \sum_{k=1}^{K} \int_{\Omega_k} \mathcal{L}\big(f(\mathbf{x};\mathbf{o}), u(\mathbf{x};\mathbf{i})\big)\, d\mathbf{x} \tag{3.18}$$

$$= \sum_{k=1}^{K} \int_{\Omega_k} \mathcal{L}\big(f(\mathbf{x};\mathbf{o}_{c(k)}), u(\mathbf{x};\mathbf{i}_{c(k)})\big)\, d\mathbf{x} \tag{3.19}$$

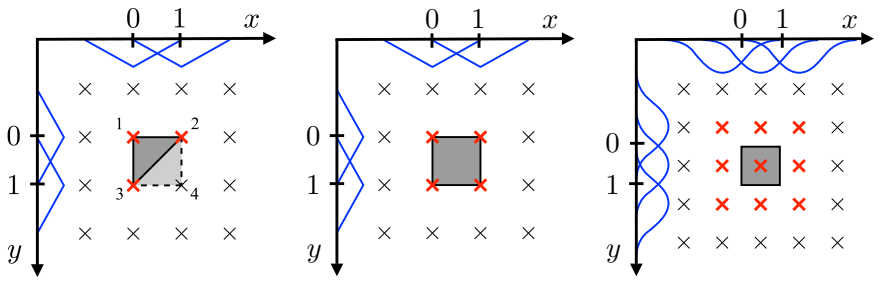
$$= \sum_{k=1}^{K} \rho(\mathbf{o}_{c(k)}, \mathbf{i}_{c(k)}). \tag{3.20}$$

This means that if we restrict ourselves to bases with local support the energy functional $\mathcal{E}(f;u)$ is equivalent to an MRF energy with cliques $\mathcal{C} = \{c(k) | k = 1, \dots, K\}$ and potential functions

$$\rho(\mathbf{o}_{c(k)}, \mathbf{i}_{c(k)}) = \int_{\Omega_k} \mathcal{L}\big(f(\mathbf{x};\mathbf{o}_{c(k)}), u(\mathbf{x};\mathbf{i}_{c(k)})\big)\, d\mathbf{x}, \tag{3.21}$$

where the cliques exert *local control* over the values on each region Ω_k. Figure 3.1 illustrates the factor graph structure of the equivalent MRF for linear, bilinear, and biquadratic FE bases. Figure 3.2 shows the connection between the integration region Ω_k and the respective cliques in the equivalent MRF.

A key characteristic of our approach lies in the rigorous realization of the *Ritz method*: We aim to eliminate the domain variable $\mathbf{x}$ in Equation (3.21) by explicit integration in closed form, rather than making discrete approximations as is frequently done (Ramani et al., 2007). We rely on FE bases of low degree for this purpose, which allow closed form integration in a number of important cases including non-quadratic regularizers. This not only yields closed form expressions for the MRF potentials in such cases, but demonstrates that a direct equivalence between large classes of variational models and MRFs can be established assuming an implementation with localized basis representations. The crux is of course, whether the integral required to eliminate the domain variable can be computed.



(a) Linear (b) Bilinear (c) Biquadratic

Figure 3.2: FE implementation with local support and MRF cliques. One local integration area Ω_k is shaded gray. The relevant control points, which also form one of the cliques of the equivalent high-order MRFs are bold red. Localized basis functions are shown in blue.

3.4 IMPLEMENTATION

To illustrate the connection between variational approaches and MRFs by use of the FEM in combination with explicit integration, we discuss three different implementations of the basic image restoration model from Equation (3.1).

3.4.1 *Linear Finite Elements*

Linear FEs rely on a triangular subdivision of the domain (Figure 3.2(a)), where each triangle supports a planar surface that interpolates the pixel values on three vertices. In contrast to Szeliski (1990a), where the data term was defined in a discrete fashion and only the spatial term was integrated, we assume that both output and input are spatially continuous and carry out an analytic integration. In more detail, for two-dimensional linear elements on triangular domains, the element functions have the form $b(x,y) = \alpha x + \beta y + \gamma$, where the coefficients α, β and γ are linear combinations of the pixels on one triangle. Specifically, we assume that the pixel values p_i are numbered according to Figure 3.2(a) and the origin is in the upper left corner, at pixel p_1. We further assume unit distance between pixels.

Then the FE representation for the upper left triangle element is given as

$$b_1(x,y;\mathbf{p}) = (p_3 - p_1)x + (p_2 - p_1)y + p_1, \tag{3.22}$$

and for the lower right triangle element

$$b_2(x,y;\mathbf{p}) = (p_4 - p_2)x + (p_4 - p_3)y + (p_2 + p_3 - p_4). \tag{3.23}$$

DATA TERM. The data term has the form

$$\int (f(x,y;\mathbf{o}) - u(x,y;\mathbf{i}))^2 \, dx dy. \tag{3.24}$$

Hence on an upper left triangle (see Figure 3.2(a)), the data term thus reduces to an evaluation of an integral of the type $\int_0^1 \int_0^{1-y} (ax + by + c)^2 \, dxdy$, which gives $\frac{1}{12}(a^2 + b^2 + 4bc + 6c^2 + ab + 4ac)$. The lower right triangle results in an integral of type $\int_0^1 \int_{1-x}^1 (ax + by + c)^2 \, dydx$, which yields $\frac{1}{12}(3a^2 + 5ab + 3b^2 + 8ac + 8bc + 6c^2)$. The data term on the square area of both triangles, which is controlled by 4 pixels, is an expression in $\hat{\mathbf{o}} = (o_1, o_2, o_3, o_4)^T$ and $\hat{\mathbf{i}} = (i_1, i_2, i_3, i_4)^T$ and is equal to the quadratic form $\frac{1}{2}(\hat{\mathbf{o}} - \hat{\mathbf{i}})^T \mathbf{M} (\hat{\mathbf{o}} - \hat{\mathbf{i}})$, with the symmetric, positive definite matrix

$$\mathbf{M} = \begin{pmatrix} 1/6 & 1/12 & 1/12 & 0 \\ 1/12 & 1/3 & 1/6 & 1/12 \\ 1/12 & 1/6 & 1/3 & 1/12 \\ 0 & 1/12 & 1/12 & 1/6 \end{pmatrix}. \tag{3.25}$$

Since the entire data term is composed of a sum of such contributions by 4 pixels, it can be written as a quadratic form $\frac{1}{2}(\mathbf{o} - \mathbf{i})^T \mathbf{A} (\mathbf{o} - \mathbf{i})$ for a sparse, symmetric, positive definite matrix $\mathbf{A}$. The matrix $\mathbf{A}$ is a non-diagonal band matrix, which reflects the pixel correlation induced by the FEM. This is in contrast to Szeliski (1990a), where $\mathbf{A}$ is diagonal (see Equation (3.10)). We conclude that the variational data term thus corresponds to a Gaussian MRF likelihood $p(\mathbf{i}|\mathbf{o})$ with correlated pixels, and show in Section 3.5 that this data term leads to increased performance.

SPATIAL TERM. Linear FEs are not classically differentiable across patch boundaries, but they admit *weak derivatives* (Evans, 2010), which are equal *a.e.* (almost everywhere) to the patch-wise defined classical derivatives. Linear FEs are particularly advantageous since any type of first-order spatial term (with an arbitrary penalty $\varphi(\cdot)$) becomes analytically integrable. This is because the partial derivatives on the triangle elements are constant *a.e.* Since both the horizontal and the vertical derivative of linear finite elements are piecewise constant *a.e.*, it follows from Equations (3.22) and (3.23) that the spatial term $E_S(\mathbf{o})$ has the form

$$\begin{aligned} 2E_S(\mathbf{o}) = &\sum_{i,j<N} \varphi\left((o_{i+1,j} - o_{i,j}, o_{i,j+1} - o_{i,j})^T\right) \\ &+ \sum_{i,j>1} \varphi\left((o_{i,j} - o_{i-1,j}, o_{i,j} - o_{i,j-1})^T\right), \end{aligned} \tag{3.26}$$

which may seem unorthodox compared to standard FDs, since the corresponding MRF energy contains a combination of forward and backward differences. Note that this does not become apparent in the previously considered quadratic case (see Equation (3.8)). It is also important to note that in contrast to Equation (3.8) the non-quadratic case results in high-order cliques; the triple cliques of the corresponding MRF prior are shown in Figure 3.1(b). These properties are a

rigorous consequence of FE calculus, and the experimental results in Section 3.5 further indicate a well-founded effect.

OPTIMIZATION. Since the main focus of this work is on bridging the gap between variational and MRF approaches to low-level vision, we focus on regularizers common in the context of variational models. Due to its convexity, the total variation (TV) regularizer enjoys particular popularity (Rudin et al., 1992; Chambolle, 2004; Pock et al., 2007; Aubert et al., 2009). We therefore choose it as main model for our implementation. The TV model penalizes discontinuities by use of the Euclidean norm $\varphi(\mathbf{x}) = \sqrt{x_1^2 + x_2^2}$. To cope with the non-differentiability at $(0,0)^T$, a common relaxation is the function $\sqrt{x_1^2 + x_2^2 + \epsilon^2}$ (Charbonnier et al., 1997). We will refer to this TV relaxation as TV + ϵ. Furthermore, we also consider an anisotropic variant of TV (+ ϵ), in which each partial derivative is penalized separately.

If we integrate TV + ϵ using linear FEs as in Equation (3.26), the MAP estimate of the resulting MRF is readily obtained by gradient descent. In the non-differentiable case of TV, an effective method for inference is much less obvious. We opt for an adaption of Chambolle's fast duality-based algorithm (Chambolle, 2004) to the FE case. The ensuing text gives an overview of the derivations. Please refer to Appendix A for further details. First, the spatial term consisting of forward and backward differences gives rise to a discrete gradient operator $\nabla : \mathbf{X} \to \mathbf{Y}$ for spaces of matrices $\mathbf{X} = \mathbb{R}^{N \times N}$ and $\mathbf{Y} = \mathbb{R}^{N \times 2N} \times \mathbb{R}^{N \times 2N}$. Intuitively, the first N columns of $\mathbf{Y}$ accommodate the forward differences of ∇ and the second N the backward differences. The divergence operator $\operatorname{div} : \mathbf{Y} \to \mathbf{X}$ is defined as the negative adjoint of the gradient operator. The FEM-induced data term leads to the Euler equation

$$\mathbf{o} \in \mathbf{A}(\mathbf{o} - \mathbf{i}) + \lambda \partial E_S(\mathbf{o}), \tag{3.27}$$

where $E_S(\mathbf{o})$ is Equation (3.26) for $\varphi(\mathbf{x}) = |\mathbf{x}|$, and ∂E_S is the *subdifferential* of E_S (Hiriart-Urruty and Lemaréchal, 1993a,b). By introducing the relation $\mathbf{v} = \frac{1}{\lambda}\mathbf{A}(\mathbf{i} - \mathbf{o})$, the solution can be obtained from minimizing

$$\mathbf{v}^T \mathbf{A}^{-1} \mathbf{v} - \mathbf{v}^T \left(\frac{2\mathbf{i}}{\lambda} \right) + \frac{2}{\lambda} E_S^*(\mathbf{v}), \tag{3.28}$$

where E_S^* denotes the *Fenchel* conjugate of E_S. The characteristic function property of E_S^* permits the solution of Equation (3.28) by optimization of the quadratic objective

$$\mathbf{v}^T \mathbf{A}^{-1} \mathbf{v} - \mathbf{v}^T \left(\frac{2\mathbf{i}}{\lambda} \right) \quad \text{subject to} \tag{3.29}$$

$$\mathbf{v} \in \{\operatorname{div}(\mathbf{p}) : \mathbf{p} \in \mathbf{Y}, |\mathbf{p}_{i,j}| \leqslant 1\}. \tag{3.30}$$

A fixed point algorithm for the completion of this task can be derived via the *Karush-Kuhn-Tucker* conditions.

3.4.2 *Bilinear Finite Elements*

Beyond linear FEs, we can also use bilinear FEs with square integration regions (see Figure 3.2(b)). Bilinear dependence on the domain variables yields the patch-wise expression

$$f(x, y; \mathbf{o}) = \sum_{k,j=0}^{1} c_{k,j} x^k y^j, \tag{3.31}$$

where the $c_{k,j}$ are linear combinations of the 2×2 clique of pixels $\mathbf{o}_{(c)}$, which exert local control (see Figure 3.2(b)). For simplicity, we assume that the coordinate system of each patch is the unit square. We note that the data term equals the patch-wise sum of integrated squared differences of bilinear functions of the form Equation (3.31). As such, the data term reduces to the straightforward, patch-wise integration of polynomials. The spatial term poses greater challenges. However, for anisotropic regularizers that penalize horizontal and vertical derivatives separately, bilinear finite elements enable exact integration in certain cases. Below, we illustrate this with two such cases. Please note that bilinear elements are not differentiable across patch boundaries, but weak derivatives exist and are equal *a.e.* to the patch-wise defined classical derivatives.

TV + ϵ CASE. For the anisotropic TV + ϵ model, the spatial term involves the patch-wise integration of $\sqrt{(\partial_x f)^2 + \epsilon^2} + \sqrt{(\partial_y f)^2 + \epsilon^2}$. The partial derivatives of the function in Equation (3.31) are linear in one domain variable (x or y). A table of integrals (Gradshteyn and Ryzhik, 2014) yields that the integral of $\sqrt{(az + b)^2 + \epsilon^2}$ over the interval $[0, 1]$ equals

$$F(a, b) = \begin{cases} \sqrt{b^2 + \epsilon^2}, & \text{if } a = 0 \\ \frac{a+b}{2a} X - \frac{b}{2a} Y + \frac{\epsilon^2}{2a} Z, & \text{otherwise} \end{cases} \tag{3.32}$$

for $X = \sqrt{(a + b)^2 + \epsilon^2}$, $Y = \sqrt{b^2 + \epsilon^2}$ and $Z = \log(X + a + b) - \log(Y + b)$. It follows that the spatial term reduces to summing over patch-wise contributions of the form $F(a, b) + F(a, c)$, where the arguments a, b, and c are the corresponding coefficients of Equation (3.31). Thus, implementing the variational model with bilinear FEs yields a MRF with 2×2 cliques, since each contribution to the energy depends on 2×2 neighboring pixels. Figure 3.1(c) depicts the corresponding factor graph. At this point, it is worthwhile to argue the differentiability of Equation (3.32). The smoothness of the integrand, together with Leibniz's rule for differentiation under the

integral, imply that the term in Equation (3.32) is infinitely differentiable. Explicit integration of the data and spatial terms thus leads to a patch-wise sum of smooth, closed-form functions in the output pixels.

LORENTZIAN CASE. To demonstrate a non-convex instance of a variational model, we stick to the previous setting, and change only the penalty function in the spatial term. A classical non-convex penalty function is the *Lorentzian* $\rho(x) = \log\left(1 + \left(\frac{x}{s}\right)^2\right)$ (Black and Rangarajan, 1996). An anisotropic, decoupled application of the Lorentzian to each partial derivative leads to the penalty term $\rho(\partial_x f) + \rho(\partial_y f)$. A table of integrals (Gradshteyn and Ryzhik, 2014) yields that the integral of $\log\left(1 + (az + b)^2\right)$ over the interval $[0, 1]$ is[5]

$$G(a, b) = \begin{cases} \log\left(1 + b^2\right), & \text{if } a = 0 \\ -2 + \log(X) + \frac{b}{a}Y + \frac{2}{a}Z, & \text{otherwise,} \end{cases} \tag{3.33}$$

for $X = 1 + (a + b)^2$, $Y = \log(X) - \log(1 + b^2)$ and $Z = \arctan(a + b) - \arctan(b)$. The spatial term is therefore composed of a sum of patch-wise contributions of the form $G(a, b) + G(a, c)$, where the arguments a, b and c are the corresponding coefficients of Equation (3.31). Again, the explicit integration of data and spatial term yields a sum of closed-form, smooth functions with local pixel dependence. Thus, here as well we obtain a MRF with 2×2 cliques. The smoothness of the term in Equation (3.33) is argued analogously to that of Equation (3.32).

3.4.3 *Biquadratic Tensor Product B-splines*

Although the experimental results clearly justify the use of linear and bilinear FEs, (bi)linear functions may still be viewed as simplistic: *E.g.*, for gray-level pixels in $[0, 255]$, the derivatives can have only limited magnitude, and furthermore, bilinear functions are not smooth across patch boundaries. Higher degree elements can address these issues. On the other hand, higher degree polynomials pose a greater challenge for explicit integration. We here consider biquadratic tensor product B-splines, which have the patch-wise form

$$f(x, y; \mathbf{o}) = \sum_{k,j=0}^{2} c_{k,j} x^k y^j, \tag{3.34}$$

where the coefficients $c_{k,j}$ are linear combinations of the 3×3 clique of pixels $\mathbf{o}_{(c)}$, which exert local control (see Figure 3.1(d) and Figure 3.2(c)).

The data term is the sum of patch-wise integrated squared differences of two expressions of the form of Equation (3.34), which depend on the input and output pixels $\mathbf{o}$ and $\mathbf{i}$. Since it only involves

5 The constant s is incorporated into the coefficients a, b for ease of notation.

polynomials, the data term reduces to a straightforward patch-wise integration. The challenge lies in the spatial term. For biquadratic B-splines in combination with the anisotropic TV + ϵ model, each partial derivative term can be evaluated similarly. Therefore, it suffices to explain our approach on one of them, say $\partial_x B$. Using the term of Equation (3.32), we note that for functions a and b, quadratic in y,

$$\int_0^1\int_0^1 \sqrt{(\partial_x B)^2 + \epsilon^2}\, dxdy = \int_0^1 F(a, b; \mathbf{o})\, dy, \tag{3.35}$$

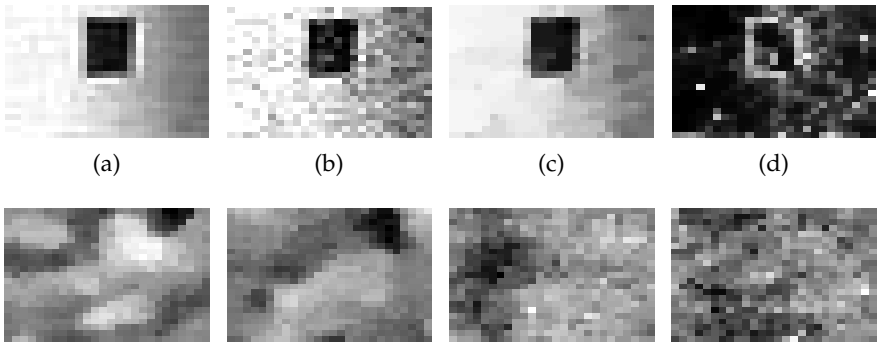
where we write $F(a, b; \mathbf{o})$ to emphasize the dependence on the pixels, as a purely notational measure. The integral in Equation (3.35) cannot be carried out in closed form, but can be approximated by a numerical quadrature scheme of the form $\sum_k F(a, b; \mathbf{o})|_{y_k} \cdot w_k$, where we again note that a and b are functions in y. In our implementation, we rely on a composite Simpson rule (Stoer and Bulirsch, 2002) as a quadrature scheme, which is of order 4. We do not claim that this choice of quadrature is optimal, but rather wish to demonstrate the feasibility of our approach using a simple method. Due to the smoothness of the term in Equation (3.32), the numerical scheme is guaranteed to converge to the true value of the integral as the step size goes to zero (Stoer and Bulirsch, 2002). Consequently, for each step size we obtain an MRF with 3×3 cliques as illustrated in Figure 3.1(d). Decreasing step sizes yield a *sequence* of MRFs of *fixed* clique size, whose potentials converge to the true potential that is equivalent to the variational formulation (due to the convergence of the quadrature). This demonstrates how the interest in higher-order element spaces can be reconciled with feasible integration. Although the integration is semi-numeric and as such of an approximative nature, the convergence is mathematically rigorous. For inference we rely on gradient descent. We call attention to the fact that the gradient $\sum_k \nabla_{\mathbf{o}} F(a, b; \mathbf{o})|_{y_k} \cdot w_k$ approximates the true gradient due to

$$\partial_{o_k} \int_0^1\int_0^1 \sqrt{(\partial_x B)^2 + \epsilon^2}\, dxdy = \int_0^1 \partial_{o_k} F(a, b; \mathbf{o})\, dy, \tag{3.36}$$

by Leibniz's rule of differentiation under the integral. We also note that the above semi-numeric technique can be carried out for the non-convex Lorentzian penalty function defined in Section 3.4.2.

3.5 EXPERIMENTATION

To demonstrate that the derived explicit connection between non-quadratic variational approaches and MRFs is not only of theoretical interest, but also provides practical advantages, we perform a series of image denoising experiments comparing standard finite difference discretizations and the MRFs obtained from explicitly integrated FEs. In both cases, the MRFs are derived from first-order vari-



(a) (b) (c) (d)

(e) $T = 0.01$ (f) $T = 1$ (g) $T = 8$ (h) $T = 20$

Figure 3.3: Metropolis sampling (Gaussian proposals, 100.000 iterations burn in) of the linear FEM, TV + ϵ model: (a) Clean input BSDS image crop. (b) Noise level $\sigma = 25$, PSNR 21.19 dB. (c) Posterior mean estimate, $T = 1$, PSNR 25.96 dB. (d) Marginal variance estimate, $T = 1$. (e–f) Prior sampling at different temperatures T.

ational models, however in case of FDs the equivalence is only approximate, whereas it is exact for our FE approach. As is typical in the literature (*e.g.*, Portilla et al., 2003), we assume i. i. d. Gaussian noise with known variance. We use and compare two regularizers: (1) the differentiable TV + ϵ penalty ($\epsilon = 0.01$, except in case of bilinear FEs where $\epsilon = 1$ is required for numerical stability), and (2) the non-differentiable TV penalty. Unless otherwise mentioned, we use an isotropic (*i.e.* gradient-magnitude based) regularizer. Restoration performance is evaluated using both PSNR and the more advanced, perceptually motivated structural similarity index (SSIM). For each permutation of method, noise level and metric, the regularization parameter λ was determined using an exhaustive grid search on 20 separate training images by maximizing the performance of the respective metric. In case of differentiable penalties, we compute the MAP estimate of the posterior MRF using conjugate gradients `[CG]`. For the MRFs derived using the non-differentiable TV model, we rely on the popular duality-based method `[DM]` of Chambolle (2004), or a variation thereof (see Sec. 3.4). Note that for the convex penalties considered here, applying inference algorithms such as belief propagation or graph cuts will likely not be beneficial. Nonetheless, the connection between variational approaches and MRFs developed here allows to apply such inference techniques to variational problems with non-convex regularizers, or to move beyond MAP estimation. We leave this for future work.

We evaluate the average image restoration performance for a variety of models and methods on 68 images (as used, *e.g.*, by Roth and Black (2005)). Table 3.1 summarizes the results, Figure 3.4 gives a visual example. We make three main observations: (1) When comparing the same variational model, the proposed FE-MRF performs at least

Table 3.1: Average restoration performance on 68 test images (from Roth and Black (2005)).

	$\sigma = 15$		$\sigma = 25$	
MRF / variational model / algorithm	PSNR	SSIM	PSNR	SSIM
FD-MRF, TV, `[DM]`	29.63	0.829	27.08	0.747
Linear FE-MRF, TV, `[DM]`	29.63	0.832	27.19	0.750
FD-MRF, TV + ϵ, `[CG]`	29.69	0.833	27.22	0.751
FD-MRF w/ FE data term, TV + ϵ, `[CG]`	29.75	0.836	27.29	0.756
FD-MRF w/ FE spatial term, TV + ϵ, `[CG]`	29.72	0.834	27.24	0.754
Linear FE-MRF, TV + ϵ, `[CG]`	**29.80**	**0.837**	**27.32**	**0.757**
FD-MRF, anisotropic TV + ϵ, `[CG]`	29.57	0.829	27.12	0.747
Bilinear FE-MRF, anisotropic TV + ϵ, `[CG]`	29.73	0.826	27.11	0.751
FoE (Roth and Black, 2005)			27.44	0.746
BLS-GSM (Portilla et al., 2003)			28.02	0.789

as well as the FD-MRF and in most cases substantially better (approximately, by 0.1 dB in PSNR). (2) The TV + ϵ penalty leads to consistently better results than standard TV regularization, which is in line with findings in the context of optical flow (Werlberger et al., 2009), likely because staircasing is avoided by a near-quadratic regularization near zero. The performance of the bilinear FE-MRF versus a corresponding FD-MRF is mixed, which may also be due to an observed numerical instability of the former. (3) By separating the contribution of the FE spatial term and the FE data term, we find that each improves upon the pure FD-MRF. The correlated data term contributes the strongest. For comparison, we also report results for two competitive denoising approaches (Portilla et al., 2003; Roth and Black, 2005), for which code is publicly available. While we note that the standard first-order variational model does not perform at the level of these more sophisticated techniques (or more recent ones (Lyu and Simoncelli, 2007; Jain and Seung, 2009)), it is simpler and enjoys widespread use to date. For completeness and to allow comparison with other methods, we also report results on a variety of standard images considered in the literature (*e.g.*, Portilla et al., 2003). Table 3.2 gives these results in terms of PSNR and SSIM.

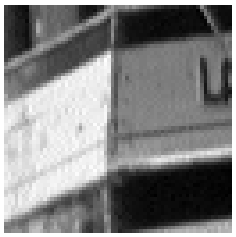
One advantage of the derived equivalence of variational models and MRFs is that we can apply probabilistic methods to variational models. Of the many possibilities, we here consider sampling. Figure 3.3 demonstrates standard Metropolis sampling from the probabilistic model induced by a variational formulation with a TV + ϵ regularizer and using the proposed linear FE-MRF. Samples from the posterior can, for example, be used to estimate the posterior mean (MMSE estimate) and the marginal variance (see Figure 3.3(c),(d)). In

Table 3.2: Restoration performance by MAP inference on the standard test images Barbara, Boat, Fingerprint, Lena, and Peppers (Portilla et al., 2003)), numbered from one to five.

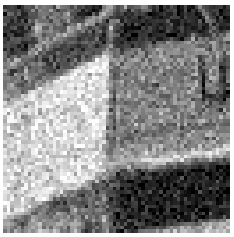
		Image-wise performance in PSNR				
σ	MRF, model, algorithm	1	2	3	4	5
15	FDM, TV, [DM]	27.90	30.47	28.17	32.28	**31.33**
	Lin. FE, TV, [DM]	27.56	30.63	28.32	32.29	30.89
	FDM, TV + ϵ, [CG]	**27.96**	30.45	28.02	32.31	31.12
	Lin. FE, TV + ϵ, [CG]	27.73	**30.71**	**28.83**	**32.53**	30.86
	FDM, aniso. TV + ϵ, [CG]	27.84	30.43	27.68	32.12	31.07
	Bilin. FE, aniso. TV + ϵ, [CG]	27.82	30.68	28.36	32.29	30.89
25	FDM, TV, [DM]	25.30	28.15	25.52	30.00	**28.59**
	Lin. FE, TV, [DM]	24.86	28.31	25.50	30.24	27.81
	FDM, TV + ϵ, [CG]	**25.46**	28.21	25.24	30.12	28.39
	Lin. FE, TV + ϵ, [CG]	25.26	**28.46**	**26.05**	**30.40**	27.88
	FDM, aniso. TV + ϵ, [CG]	25.37	28.18	24.95	29.91	28.38
	Bilin. FE, aniso. TV + ϵ, [CG]	25.16	28.42	25.63	30.19	27.98

		Image-wise performance in SSIM				
σ	MRF, model, algorithm	1	2	3	4	5
15	FDM, TV, [DM]	0.812	0.814	0.917	0.858	**0.882**
	Lin. FE, TV, [DM]	0.803	0.819	0.923	**0.859**	0.878
	FDM, TV + ϵ, [CG]	**0.822**	0.814	0.917	0.855	0.876
	Lin. FE, TV + ϵ, [CG]	0.809	**0.820**	**0.930**	**0.859**	0.880
	FDM, aniso. TV + ϵ, [CG]	0.813	0.811	0.910	0.852	0.875
	Bilin. FE, aniso. TV + ϵ, [CG]	0.806	0.819	0.923	0.858	0.880
25	FDM, TV, [DM]	0.720	0.748	0.865	0.796	0.820
	Lin. FE, TV, [DM]	0.706	0.754	0.861	**0.812**	**0.835**
	FDM, TV + ϵ, [CG]	**0.727**	0.747	0.848	0.810	0.829
	Lin. FE, TV + ϵ, [CG]	0.720	**0.756**	**0.874**	**0.812**	0.831
	FDM, aniso. TV + ϵ, [CG]	0.719	0.743	0.840	0.803	0.823
	Bilin. FE, aniso. TV + ϵ, [CG]	0.713	**0.756**	0.865	0.807	0.828

(a) Full size image



(b) Image detail



(c) Noisy image, noise level $\sigma = 25$

(d) FDM-MRF, TV + ϵ, `[CG]`

(e) Linear FE-MRF, TV + ϵ, `[CG]`

Figure 3.4: Image denoising with a non-quadratic variational model and the corresponding MRF realizations. The FE-MRF result exhibits fewer speckle artifacts (best viewed on screen). The image is from the USC-SIPI database (Weber, 2018).

the future, posterior samples may also be used to infer MAP estimates of non-convex regularizers via simulated annealing. Figure 3.3(e)-(h) shows how sampling the equivalent MRF can also give insights into the generative properties of the underlying variational model.

3.6 CONCLUSIONS

In this chapter we investigated the connections of modern, non-quadratic first-order variational models and Markov random fields. Based on finite elements and explicit integration, we derived localized potential functions of an equivalent MRF, which provided a rigorous connection between the two approaches. We demonstrated the feasibility of the approach with several examples, and adapted a duality-based inference algorithm to MAP inference in MRFs derived from FE implementations of TV-based variational models. Moreover, we gave an experimental analysis of the derived connection in an image denoising application, where we found improved performance compared to standard discretization schemes for variational models. Finally, we illustrated the connection to probabilistic models by sampling the equivalent MRF prior and posterior. Future work should consider applying probabilistic learning and inference to the MRFs equivalent to the variational formulation, such as learning the penalty function, performing MAP estimation for non-convex penalties, or carrying out MMSE estimation, *e.g.* by variational inference. In addition, the estab-

lished connection allows to investigate the generative properties of variational models. Future work may also consider second-order priors and other problem domains, such as correspondence problems or segmentation.

4

BAYESIAN DEBLURRING

CONTENTS

Conventional non-blind image deblurring methods involve natural image priors and maximum a-posteriori (MAP) estimation. As a consequence of MAP estimation, separate pre-processing steps such as training of the regularization parameter are necessary to avoid user interaction. Moreover, MAP estimates involving standard natural image priors have been found lacking in terms of restoration performance. To address these issues, we introduce a Bayesian framework for non-blind deblurring. Thereby, a sampling-based technique allows to perform deblurring using the Bayesian minimum mean squared error (MMSE) estimate, which requires no regularization parameter and yields higher performance than MAP estimates when combined with a learned high-order image prior. Another advantage of the proposed framework is that it permits integrated estimation of the unknown noise strength. A quantitative and qualitative evaluation demonstrates favorable deblurring results.

4.1 INTRODUCTION

Although image blur is sometimes used for artistic purposes, it frequently corrupts valuable image information and produces visually disturbing images. Deblurring techniques thus attempt to restore a sharp explanation from a blurred input image. This chapter is concerned with *non-blind* deblurring, where the blur is assumed to be known, which is an important component of the more general blind deblurring problem (Levin et al., 2009). Even when the image blur is known, for example from inertial sensors (Joshi et al., 2010), deblurring is a difficult problem, partly rooted in the loss of image information due to the presence of zeros in the kernel spectrum, as occurs,

e.g., in the case of motion and defocus blur. This is further exacerbated by image noise, which arises from the image capturing process. The simplest conceivable approach of deblurring by inverting the blur matrix cannot effectively overcome these difficulties (see Section 1.3 for a more extensive discussion of these challenges).

Instead, most techniques rely on standard natural image priors based on Markov random fields (MRFs) to regularize the problem and to avoid restoration artifacts. The restored image is coupled to the blurred input through a likelihood model, which requires the image noise level to be known. To compute the deblurred image, most approaches rely on MAP estimation (*e.g.*, Levin et al., 2007; Krishnan and Fergus, 2009). However, such conventional MAP approaches have shortcomings: *(1)* For best results, the influence of prior and likelihood needs to be calibrated with a regularization parameter. This parameter depends on the noise magnitude; therefore, a set of suitable regularization parameters must be determined in an off-line training step for a necessarily limited set of noise levels, which must be given by the user or estimated in a separate process prior to deblurring. *(2)* Conventional MAP estimates have been found to either exhibit oversmoothing or residual high-frequency artifacts (Cho et al., 2010; Schmidt et al., 2010).

In the non-blind image recovery procedure presented in this chapter, we replace MAP with MMSE estimation based on Gibbs sampling (Schmidt et al., 2010), which reduces oversmoothing and allows to eliminate the regularization parameter when combined with an appropriate image prior. Furthermore, we employ a learned high-order MRF prior (Roth and Black, 2009) for regularization. Another advantage of our Bayesian approach is that unknown parameters, such as the noise variable, can be eliminated by approximately integrating over them using Markov chain Monte Carlo (MCMC) sampling.

We evaluate our non-blind deblurring approach experimentally and find both quantitatively and qualitatively favorable performance in comparison to other popular methods.

4.2 RELATED WORK

This section is designed to present a selection of related publications pertaining directly to the subject matter at hand. To understand how the research enumerated below links into a larger context of deblurring literature, we refer the reader to Section 2.3.

The practical significance of non-blind deblurring not only includes applications with known blur kernel (*e.g.*, Levin et al., 2007), but also stems from the design of state-of-the-art blind deblurring algorithms. Levin et al. (2009) theoretically and practically demonstrate the stability of *first* estimating the kernel from a marginalized density, and *then* inferring a sharp image by non-blind deblurring with the kernel

estimate. Although some algorithms differ from this and alternate between kernel and latent image estimation (*e.g.*, Shan et al., 2008), non-blind deblurring forms an important element of not only state-of-the-art image-based blind methods (Fergus et al., 2006; Yuan et al., 2007; Joshi et al., 2008; Shan et al., 2008; Cho and Lee, 2009), but also of highly competitive hardware-based techniques (Ben-Ezra and Nayar, 2004; Tai et al., 2008; Joshi et al., 2010).

A classical approach to the problem of non-blind deblurring is the Lucy-Richardson method (Richardson, 1972; Lucy, 1974), which is known to cause ringing artifacts, but frequently is used as a baseline (Shan et al., 2008; Krishnan and Fergus, 2009). Ringing can be reduced if a pair of blurred and noisy images is available (Yuan et al., 2007). More recent deblurring approaches involve natural image priors and MAP estimation. Krishnan and Fergus (2009) propose a fast MAP algorithm with a first-order prior by exploiting a half-quadratic scheme (Geman and Yang, 1995). Shan et al. (2008) use a likelihood model based on derivatives of up to second order and a two-level prior. Levin et al. (2007) propose a MAP algorithm with a high-order prior involving second-order derivative filters. This method is widely used due to its leading performance (Cho and Lee, 2009; Levin et al., 2009; Joshi et al., 2010), or is often compared to experimentally (Shan et al., 2008; Krishnan and Fergus, 2009).

As discussed above, standard MAP approaches to deblurring have a number of disadvantages. On one hand, conventional MAP estimates have been found to yield either oversmoothed results lacking textural detail, or results with high-frequency artifacts. Cho et al. (2010) address this issue with a content-aware prior, which adapts the image model locally to the properties of the respective image region. Schmidt et al. (2010) replace MAP with MMSE estimation, and show improved results for image denoising. We here follow the approach of Schmidt et al. (2010) and generalize it to non-blind deblurring. Furthermore, the sampling-based approach lends itself to the task of noise level estimation.

By contrast, conventional MAP methods typically rely on the specification of the noise magnitude or regularization weight, which may be pre-determined for the case of a fixed camera setup or estimated as an additional processing step. One previous approach is to use variational inference, which involves fitting an easier to handle probability to the posterior. The approximate distribution permits to analytically derive marginal expected values of unknown variables such as the hidden image and noise magnitude. Particularly, the variational method of Miskin and MacKay (2000) for blind deblurring has a simultaneous noise estimation component, but the assumption of i. i. d. pixels is a limiting factor, and experiments are restricted to cartoon images. The variational framework is also used by Fergus et al. (2006) for blur kernel estimation. However, the noise is not estimated in the

spatial, but in the gradient domain, which is equipped with a conventional prior. This stands in contrast to the MCMC-based non-blind deconvolution presented in this chapter, which permits high-order Field of Experts (FOE) priors and spatial noise estimation.

The difficulty of specifying viable regularization parameters is not exclusive to deblurring. For stereo vision, Zhang and Seitz (2007) propose simultaneous depth field and MRF parameter estimation. In the context of optical flow computation, Krajsek and Mester (2006) approximately marginalize over the displacement field using Laplace's method, which permits them to maximize the marginal likelihood w.r.t. the model parameters. By contrast, our approach relies on Gibbs sampling, which is typically more accurate as approximate inference procedure.

4.3 DEBLURRING WITH HIGH-ORDER PRIORS

As is usual in the deblurring literature, we assume that the unknown sharp image $\mathbf{x} \in \mathbb{R}^m$ is blurred with a blur matrix $\mathbf{K} \in \mathbb{R}^{n \times m}$ and corrupted with additive white Gaussian noise $\mathbf{n}$:

$$\mathbf{y} = \mathbf{K}\mathbf{x} + \mathbf{n}, \qquad \mathbf{n} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}). \tag{4.1}$$

Here, $\mathbf{y} \in \mathbb{R}^n$ is the observed, blurred image. We note that all subsequent derivations also hold for the case of spatially varying, non-uniform blur. However, since there is little ground-truth data available for spatially varying blur, our experiments are limited to spatially uniform blur. In other terms, we usually assume that $\mathbf{K}\mathbf{x} \equiv \mathbf{k} \otimes \mathbf{x}$ corresponds to a convolution of the desired image $\mathbf{x}$ with the blur kernel $\mathbf{k}$.

In this chapter, we consider the problem of non-blind image deblurring with a known blur matrix $\mathbf{K}$. Taking a Bayesian approach thus amounts to formulating the posterior

$$p(\mathbf{x}|\mathbf{y}, \mathbf{K}, \sigma) \propto p(\mathbf{y}|\mathbf{x}, \mathbf{K}, \sigma) \cdot p(\mathbf{x}). \tag{4.2}$$

The assumption of additive white Gaussian noise gives rise to the likelihood

$$p(\mathbf{y}|\mathbf{x}, \mathbf{K}, \sigma) = \mathcal{N}(\mathbf{y}; \mathbf{K}\mathbf{x}, \sigma^2 \mathbf{I}). \tag{4.3}$$

The term $p(\mathbf{x})$ denotes a natural image prior. This prior is necessary, since simply inverting the blur matrix $\mathbf{K}$ to recover the original image $\mathbf{x}$ is infeasible due to the presence of noise or a misspecified blur kernel. Hence, deblurring techniques employ sparse image priors to regularize the solution to this ill-posed problem (*e.g.*, Levin et al., 2007; Krishnan and Fergus, 2009).

HIGH-ORDER MRF PRIOR. Instead of the hand-defined and frequently gradient-based priors prevalent in deblurring, we here rely on a learned high-order prior, specifically Fields of Experts (FOEs) (Roth and Black, 2009). These high-order MRFs are based on spatially extended cliques, and capture the properties of a natural image $\mathbf{x}$ in terms of responses to a bank of learned, linear filters:

$$p(\mathbf{x}) = \frac{1}{Z} e^{-\epsilon \|\mathbf{x}\|^2/2} \prod_{c \in \mathcal{C}} \prod_{i=1}^{N} \phi(\mathbf{J}_i^T \mathbf{x}_{(c)}; \boldsymbol{\alpha}_i) \tag{4.4}$$

Here, the response to the learned, linear filters $\mathbf{J}_i$ is modeled as the product of expert functions ϕ with parameters $\boldsymbol{\alpha}_i$, which are learned as well. Moreover, the MRF cliques are denoted as $c \in \mathcal{C}$, and Z denotes the partition function. The broad Gaussian factor $e^{-\epsilon \|\mathbf{x}\|^2/2}$ ensures that the model is normalizable.

Instead of the original Student-t experts (Roth and Black, 2009), we follow Schmidt et al. (2010) and model the expert functions using Gaussian scale mixtures (GSMs) (Wainwright and Simoncelli, 2000) as

$$\phi(\mathbf{J}_i^T \mathbf{x}_{(c)}; \boldsymbol{\alpha}_i) = \sum_{j=1}^{J} \alpha_{ij} \mathcal{N}(\mathbf{J}_i^T \mathbf{x}_{(c)}; 0, \eta_{ij}^2). \tag{4.5}$$

Apart from being a more faithful model of natural images, this formulation admits efficient Gibbs sampling-based inference (Levi, 2009), which we later exploit for our integrated noise estimation approach. To that end, the FOE density from Equation (4.4) can be augmented with discrete latent variables $\mathbf{z}$ to yield the joint distribution

$$p(\mathbf{x}, \mathbf{z}) \propto e^{-\epsilon \|\mathbf{x}\|^2/2} \prod_{c \in \mathcal{C}} \prod_{i=1}^{N} \alpha_{i z_{ic}} \mathcal{N}(\mathbf{J}_i^T \mathbf{x}_{(c)}; 0, \eta_{i z_{ic}}^2), \tag{4.6}$$

from which the FOE model in Equation (4.4) arises by marginalizing over the latent variables $\mathbf{z}$.

The advantage of the augmented distribution $p(\mathbf{x}, \mathbf{z})$ is that the conditional distributions are tractable. Specifically, $p(\mathbf{x}|\mathbf{z})$ is a multivariate Gaussian and $p(\mathbf{z}|\mathbf{x})$ is a discrete distribution (Schmidt et al., 2010). To verify these claims in higher detail, we first remark that

$$\sum_{i=1}^{N} \sum_{c \in \mathcal{C}} -\frac{(\mathbf{J}_i^T \mathbf{x}_{(c)})^2}{2\eta_{i z_{ic}}^2} = -\frac{1}{2} \mathbf{x}^T \left(\sum_{i=1}^{N} \mathbf{T}_{\mathbf{J}_i} \mathbf{Z}_i \mathbf{T}_{\mathbf{J}_i}^T \right) \mathbf{x}, \tag{4.7}$$

whereby $\mathbf{T}_{\mathbf{J}_i}$ denotes the Toeplitz matrix expressing convolution by the filter $\mathbf{J}_i$, and $\mathbf{Z}_i$ is the diagonal matrix with entries $\left\{\eta_{iz_{ic}}^{-2}, c \in \mathcal{C}\right\}$. Using the identity from Equation 4.7, we obtain

$$
\begin{aligned}
p(\mathbf{x}|\mathbf{z}) &\propto p(\mathbf{x}, \mathbf{z}) && (4.8)\\
&\propto e^{-\epsilon\|\mathbf{x}\|^2/2} \prod_{c\in\mathcal{C}} \prod_{i=1}^{N} \mathcal{N}\left(\mathbf{J}_i^T \mathbf{x}_{(c)}; 0, \eta_{iz_{ic}}^2\right) && (4.9)\\
&\propto e^{-\epsilon\|\mathbf{x}\|^2/2} \prod_{c\in\mathcal{C}} \prod_{i=1}^{N} \exp\left(-\frac{\left(\mathbf{J}_i^T \mathbf{x}_{(c)}\right)^2}{2\eta_{iz_{ic}}^2}\right) && (4.10)\\
&\propto \exp\left(-\frac{\epsilon\|\mathbf{x}\|^2}{2} + \sum_{i=1}^{N} \sum_{c\in\mathcal{C}} -\frac{\left(\mathbf{J}_i^T \mathbf{x}_{(c)}\right)^2}{2\eta_{iz_{ic}}^2}\right) && (4.11)\\
&= \exp\left(-\frac{1}{2}\mathbf{x}^T \left(\epsilon\mathbf{I} + \sum_{i=1}^{N} \mathbf{T}_{\mathbf{J}_i} \mathbf{Z}_i \mathbf{T}_{\mathbf{J}_i}^T\right) \mathbf{x}\right) && (4.12)\\
&\propto \mathcal{N}\left(\mathbf{x}; 0, \mathbf{P}_{\mathbf{z}}^{-1}\right), && (4.13)
\end{aligned}
$$

with precision matrix $\mathbf{P}_{\mathbf{z}} = \epsilon\mathbf{I} + \sum_{i=1}^{N} \mathbf{T}_{\mathbf{J}_i} \mathbf{Z}_i \mathbf{T}_{\mathbf{J}_i}^T$, see Schmidt (2010). The matrix $\mathbf{P}_{\mathbf{z}}$ is the same as in denoising and depends on the linear filters $\mathbf{J}_i$ and the current value of the auxiliary variables $\mathbf{z}$. Furthermore, we observe that given the image $\mathbf{x}$, the discrete indicator variables z_{ic} are conditionally independent and distributed according to

$$
p(z_{ic}|\mathbf{x}) \propto \alpha_{iz_{ic}} \mathcal{N}(\mathbf{J}_i^T \mathbf{x}_{(c)}; 0, \eta_{iz_{ic}}^2), \tag{4.14}
$$

an uncomplicated expression to sample from.

4.4 BAYESIAN DEBLURRING USING SAMPLING

To infer the deblurred image from the posterior in Equation (4.2), we extend the approach of Schmidt et al. (2010) and in contrast to previous deblurring methods compute the posterior mean, or Bayesian minimum mean squared error (MMSE) estimate:

$$
\begin{aligned}
\hat{\mathbf{x}} &= \arg\min_{\tilde{\mathbf{x}}} \int \|\tilde{\mathbf{x}} - \mathbf{x}\|^2 p(\mathbf{x}|\mathbf{y}, \mathbf{K}, \sigma)\, d\mathbf{x} && (4.15)\\
&= \arg\min_{\tilde{\mathbf{x}}} \int \left(\|\tilde{\mathbf{x}}\|^2 - 2\tilde{\mathbf{x}}^T \mathbf{x} + \|\mathbf{x}\|^2\right) p(\mathbf{x}|\mathbf{y}, \mathbf{K}, \sigma)\, d\mathbf{x} && (4.16)\\
&= \arg\min_{\tilde{\mathbf{x}}} \left(\|\tilde{\mathbf{x}}\|^2 - 2\tilde{\mathbf{x}}^T \mathrm{E}[\mathbf{x}|\mathbf{y}, \mathbf{K}, \sigma]\right) && (4.17)\\
&= \mathrm{E}[\mathbf{x}|\mathbf{y}, \mathbf{K}, \sigma]. && (4.18)
\end{aligned}
$$

The advantage over the more common MAP approach, at least for image denoising (Schmidt et al., 2010), is that that it leads to superior results, both for smooth and textured image regions. Secondly, MMSE estimates yield a higher correlation between the image restoration

performance and the generative quality of the model. This on one hand lets us take advantage of powerful learned priors, and on the other hand allows us to work without any regularization parameter that balances the prior and the likelihood, which is very desirable especially when the noise level is not known.

We perform MMSE estimation by extending the sampling approach of Schmidt et al. (2010) to image deblurring. To that end, we note that the posterior can be augmented with discrete latent variables $\mathbf{z}$:

$$p(\mathbf{x}, \mathbf{z}|\mathbf{y}, \mathbf{K}, \sigma) \propto p(\mathbf{y}|\mathbf{x}, \mathbf{K}, \sigma) \cdot p(\mathbf{x}, \mathbf{z}). \tag{4.19}$$

Due to the Gaussian form of the likelihood from Equation (4.3) we obtain the conditional distributions

$$p(\mathbf{z}|\mathbf{x}, \mathbf{y}, \mathbf{K}, \sigma) \propto \prod_{c \in \mathcal{C}} \prod_{i=1}^{N} \alpha_{iz_{ic}} \mathcal{N}\left(\mathbf{J}_i^T \mathbf{x}_{(c)}; 0, \eta^2_{iz_{ic}}\right) \tag{4.20}$$

$$p(\mathbf{x}|\mathbf{z}, \mathbf{y}, \mathbf{K}, \sigma) \propto \mathcal{N}\left(\mathbf{y}; \mathbf{K}\mathbf{x}, \sigma^2 \mathbf{I}\right) \cdot \mathcal{N}\left(\mathbf{x}; 0, \mathbf{P}_\mathbf{z}^{-1}\right) \tag{4.21}$$

$$\propto \mathcal{N}\left(\mathbf{x}; \mathbf{P}_\mathbf{K}^{-1}\left(\frac{\mathbf{K}^T\mathbf{y}}{\sigma^2}\right), \mathbf{P}_\mathbf{K}^{-1}\right) \cdot \mathcal{N}\left(\mathbf{x}; 0, \mathbf{P}_\mathbf{z}^{-1}\right) \tag{4.22}$$

$$\propto \mathcal{N}\left(\mathbf{x}; \mathbf{Q}_\mathbf{z}^{-1}\left(\frac{\mathbf{K}^T\mathbf{y}}{\sigma^2}\right), \mathbf{Q}_\mathbf{z}^{-1}\right) \tag{4.23}$$

using the notation $\mathbf{P}_\mathbf{K} = \frac{1}{\sigma^2}\mathbf{K}^T\mathbf{K}$ and $\mathbf{Q}_\mathbf{z} = \mathbf{P}_\mathbf{z} + \mathbf{P}_\mathbf{K}$. Equation (4.23) follows from a standard rule for the product of two Gaussians. Note that the conditional $p(\mathbf{z}|\mathbf{x}, \mathbf{y}, \mathbf{K}, \sigma) = p(\mathbf{z}|\mathbf{x})$ is not affected by the likelihood term and hence is the same as in the denoising case (Schmidt et al., 2010). Sampling from Equation (4.20) is easy, because the distribution decomposes and each z_{ic} can be sampled independently from a univariate discrete distribution.

The conditional distribution $p(\mathbf{x}|\mathbf{z}, \mathbf{y}, \mathbf{K}, \sigma)$, on the other hand, is a generalization of the denoising case (where $\mathbf{K} = \mathbf{I}$). Equation (4.21) is a Gaussian in $\mathbf{x}$ with precision (inverse covariance) matrix $\mathbf{Q}_\mathbf{z} = \mathbf{P}_\mathbf{z} + \mathbf{P}_\mathbf{K}$ and mean $\boldsymbol{\mu} = \mathbf{Q}_\mathbf{z}^{-1}\mathbf{K}^T\frac{\mathbf{y}}{\sigma^2}$, and can be sampled by solving two large sparse systems of linear equations (Levi, 2009; Schmidt, 2010), one of which to determine the mean $\boldsymbol{\mu}$. We outline this scheme in the following. To begin with, $\mathbf{Q}_\mathbf{z}$ can be written as the matrix product $\mathbf{V}\mathbf{Z}\mathbf{V}^T$, where

$$\mathbf{V} = \begin{bmatrix} \mathbf{T}_{J_1} & \dots & \mathbf{T}_{J_N} & \mathbf{I} & \mathbf{K}^T \end{bmatrix}, \tag{4.24}$$

$$\mathbf{Z} = \begin{bmatrix} \mathbf{Z}_1 & 0 & \dots & & 0 \\ 0 & \mathbf{Z}_2 & & & \vdots \\ \vdots & & \ddots & & \\ & & & \mathbf{Z}_N & \\ & & & & \epsilon \cdot \mathbf{I} \\ 0 & \dots & & & \mathbf{I}/\sigma^2 \end{bmatrix}. \tag{4.25}$$

A sample $\hat{\mathbf{x}}^{(t)}$ from $\mathcal{N}(\mathbf{0}, \mathbf{Q}_{\mathbf{z}}^{-1})$ is generated by solving

$$\mathbf{V}\mathbf{Z}\mathbf{V}^{T}\hat{\mathbf{x}}^{(t)} = \mathbf{V}\sqrt{\mathbf{Z}}\mathbf{s}, \quad \mathbf{s} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}). \tag{4.26}$$

Finally, we obtain a sample $\mathbf{x}^{(t)}$ from $p(\mathbf{x}|\mathbf{z}, \mathbf{y}, \mathbf{K}, \sigma)$ as $\boldsymbol{\mu} + \hat{\mathbf{x}}^{(t)}$.

Sampling from the posterior (Equation (4.2)) thus proceeds using a Gibbs sampler that alternates between sampling from Equations (4.20) and (4.21) to obtain a sequence of samples $\{\{\mathbf{z}^{(1)}, \mathbf{x}^{(1)}\}, \ldots, \{\mathbf{z}^{(T)}, \mathbf{x}^{(T)}\}\}$. The MMSE estimate of $\mathbf{x}$ is approximated by averaging samples $\mathbf{x}^{(t)}$ from the posterior after B burn-in iterations ($B < t \leqslant T$); the samples of $\mathbf{z}$ are simply discarded at the end. Alternatively, we can use a Rao-Blackwellized MMSE estimator of $\mathbf{x}$ by averaging the conditional expectations from Equation (4.21) (see Papandreou and Yuille, 2010):

$$\hat{\mathbf{x}}_{\mathrm{RB}} \approx \frac{1}{T-B} \sum_{t=B+1}^{T} \mathbf{Q}_{\mathbf{z}^{(t)}}^{-1} \mathbf{K}^{T} \frac{\mathbf{y}}{\sigma^2}. \tag{4.27}$$

We find that essentially the same performance can be achieved using either estimator, although many fewer iterations of the Gibbs sampler are necessary to satisfy our conservative convergence criteria (similar to Schmidt et al. (2010)) when using Rao-Blackwellization. Hence, all results in the remainder of the chapter were obtained using this approach.

4.5 INTEGRATED NOISE ESTIMATION[1]

MAP solutions in low-level vision typically require a regularization variable λ to properly adjust the weighting of prior and likelihood components factoring into the posterior. This variable must be altered with the noise magnitude to obtain optimal performance. In practice, a look-up table or function mapping from noise level to regularization parameter is often generated prior to method application. In the conventional MAP setting, the user is expected to specify an estimate of the noise level, whereby any error in selection can have substantial ramifications on algorithm performance. On this subject, Figure 4.1 depicts the dependence of deblurring results on the choice of noise magnitude; it is evident that an accurate specification of noise level is paramount to achieving the best results of any displayed deblurring procedure. A distinctive feature of the MMSE-oriented deblurring method which we propose in Section 4.4 is its independence of any regularization variables determined from prior offline training. All other procedures in Figure 4.1, on the other hand, need to be instantiated with a favorable choice of regularizing parameter.

In this section, we expand the approach of Section 4.4 by integrated noise estimation. In more detail, we follow the Bayesian paradigm

1 This summary of the noise estimation approach of Schmidt (2016) is given for completeness to illustrate another advantage of the Bayesian approach to deblurring. It is not an original contribution of the present work.

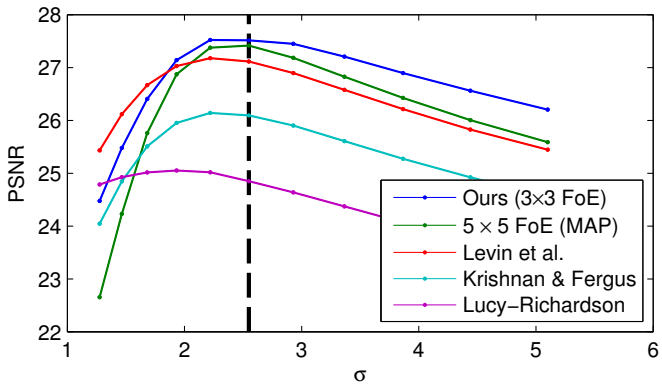


Figure 4.1: Dependence of deblurring results on the noise level. Average results for 8 images and various assumed noise levels (ground truth $\sigma_{GT} = 2.55$). All methods suffer from an incorrect choice of σ. Especially for MAP-based approaches the best performance is achieved for $\sigma < \sigma_{GT}$. The trained regularization parameter is thus not fully representative of the test set.

and eliminate the noise standard deviation σ by simply integrating over it (approximately):

$$p(\mathbf{x}|\mathbf{y}, \mathbf{K}) = \int p(\mathbf{x}, \sigma|\mathbf{y}, \mathbf{K}) \, d\sigma. \tag{4.28}$$

For this purpose, we incorporate the noise variable σ into an expanded joint probability $p(\mathbf{x}, \mathbf{z}, \sigma|\mathbf{y}, \mathbf{K})$. Note that in practice, we find it sufficient to employ a non-informative prior $p(\sigma) = \text{const.}$ on the noise level, as the corrupted/hidden image pair and blur kernel exert powerful enough constraints to guide the noise estimation. To execute the approximate integration of Equation (4.28), we expand the Gibbs sampling scheme delineated in Section 4.4 by additionally sampling from $p(\sigma|\mathbf{x}, \mathbf{z}, \mathbf{y}, \mathbf{K})$, the conditional probability for the noise level. Following Fergus et al. (2006), the latter probability may be reformulated as a Gamma distribution $\mathcal{G}(x; a, b) = \frac{x^{a-1} e^{-x/b}}{b^a \Gamma(a)}$ on the noise precision,

$$\begin{aligned} p(\sigma|\mathbf{x}, \mathbf{z}, \mathbf{y}, \mathbf{K}) &\propto \mathcal{N}(\mathbf{y}; \mathbf{K}\mathbf{x}, \sigma^2 \mathbf{I}) && (4.29) \\ &\propto \sigma^{-n} \exp\left(-\frac{\|\mathbf{y} - \mathbf{K}\mathbf{x}\|^2}{2\sigma^2}\right) && (4.30) \\ &\propto \mathcal{G}\left(\frac{1}{\sigma^2}; \frac{n}{2} + 1, \frac{2}{\|\mathbf{y} - \mathbf{K}\mathbf{x}\|^2}\right). && (4.31) \end{aligned}$$

The Gibbs sampling pattern alternates between σ, $\mathbf{z}$ and $\mathbf{x}$. Thus, the noise level must not be known in advance. Rather, it is implicitly estimated in the following MMSE solution for the deblurred image:

$$\hat{\mathbf{x}} = \arg\min_{\tilde{\mathbf{x}}} \iint \|\tilde{\mathbf{x}} - \mathbf{x}\|^2 p(\mathbf{x}, \sigma|\mathbf{y}, \mathbf{K}) \, d\mathbf{x} \, d\sigma = \mathrm{E}[\mathbf{x}|\mathbf{y}, \mathbf{K}]. \tag{4.32}$$

In practice, we compute this result by replacing the ground-truth noise level in Equation (4.27) with the corresponding sample $\sigma^{(t)}$.

If we wish to estimate the noise magnitude σ, it is equally viable to approximate the MMSE solution

$$\hat{\sigma} = \arg\min_{\tilde{\sigma}} \iint \|\tilde{\sigma} - \sigma\|^2 p(\mathbf{x}, \sigma|\mathbf{y}, \mathbf{K}) \, d\mathbf{x} \, d\sigma = E[\sigma|\mathbf{y}, \mathbf{K}] \tag{4.33}$$

by the average of noise level samples $\sigma^{(t)}$. Another option is to compute the Rao-Blackwellized MMSE solution as

$$\hat{\sigma}_{RB} \approx \frac{1}{T - B} \sum_{t=B+1}^{T} \left(\frac{n+2}{\|\mathbf{y} - \mathbf{K}\mathbf{x}^{(t-1)}\|^2} \right)^{-1/2}, \tag{4.34}$$

which is derived from the conditional means of Equation (4.31).

In comparison to the estimation method of Fergus et al. (2006), our MCMC approach extends effortlessly to high-order image priors and determines the noise magnitude in the spatial domain. Further, in contrast to conventional MAP methods, we integrate our noise estimation technique into the deblurring algorithm by formulating the noise level as an additional posterior variable and marginalizing over it. The benefit is that noise estimation is automatically tailored to the given restoration scenario, in this case image deconvolution.

A further interesting facet occurs when the blur reduces to the identity $\mathbf{K} = \mathbf{I}$. In this case, sampling from $p(\mathbf{x}, \mathbf{z}, \sigma|\mathbf{y}, \mathbf{K})$ becomes a method for *blind* denoising (without knowledge of the noise level).

4.6 EXPERIMENTS

In this section, we evaluate our approach for non-blind deblurring experimentally[2]. To separate the contribution of our general approach from the effects of the learned, high-order prior, we also report results for a learned pairwise MRF. If unspecified, the experimental discussion below refers to using the 3×3 FOE. To facilitate comparisons, we use the parameters of the learned pairwise MRF and 3×3 FOE models from Schmidt et al. (2010).

NON-BLIND DEBLURRING. A general problem with the evaluation of non-blind deblurring algorithms is the scarcity of ground-truth data or even realistic blur kernels. We use the eight publicly available blur kernels from Levin et al. (2009) on eight different images of size 128×128 pixels each, to synthetically blur 64 images overall; the images were randomly cropped from the denoising test set of 68 images proposed by Roth and Black (2009). After blurring the images, we added white Gaussian noise of varying strength to obtain three test sets for noise levels 1%, 3%, and 5% ($\sigma = 2.55, 7.65, 12.75$). To simulate realistic conditions, all pixel values were subsequently rounded

2 MATLAB code is available at http://www.visinf.tu-darmstadt.de/media/visinf/software/deblurring_demo-10.zip

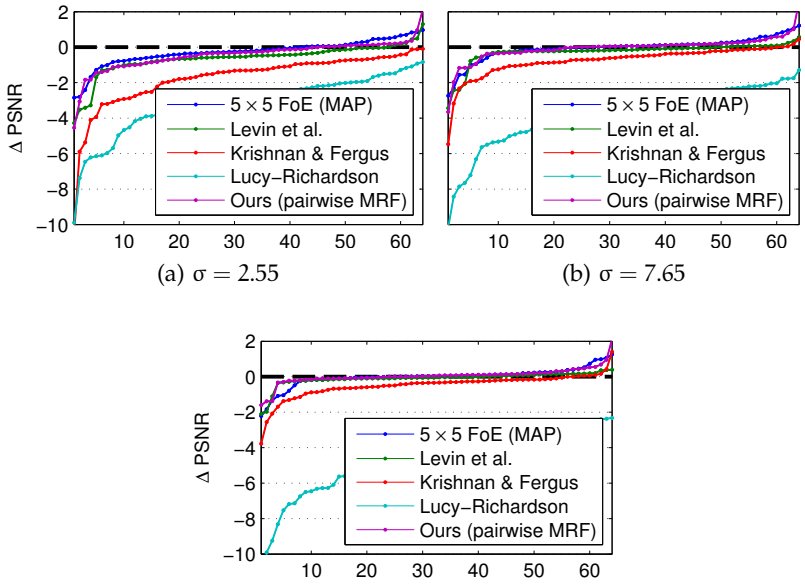
	Results in PSNR (dB)					
	$\sigma = 2.55$		$\sigma = 7.65$		$\sigma = 12.75$	
Method	GT	NE	GT	NE	GT	NE
Richardson (1972); Lucy (1974)	25.38	25.34	21.85	21.88	19.83	19.86
Krishnan and Fergus (2009)	26.97	26.86	24.91	24.88	23.93	23.94
Levin et al. (2007)	28.03	27.96	25.36	25.36	24.29	24.34
5×5 FOE (MAP) (Roth and Black, 2009)	28.44	28.33	25.66	25.59	24.48	24.43
Ours (pairwise MRF)	28.24	28.17	25.63	25.58	**24.51**	**24.48**
Ours (3×3 FOE)	**28.66**	**28.61**	**25.68**	**25.64**	24.46	24.43

	Results in SSIM					
	$\sigma = 2.55$		$\sigma = 7.65$		$\sigma = 12.75$	
Method	GT	NE	GT	NE	GT	NE
Richardson (1972); Lucy (1974)	0.703	0.697	0.423	0.425	0.244	0.245
Krishnan and Fergus (2009)	0.800	0.793	0.671	0.669	0.608	0.605
Levin et al. (2007)	0.823	0.817	0.689	0.686	0.625	0.624
5×5 FOE (MAP) (Roth and Black, 2009)	0.842	0.835	**0.711**	**0.708**	**0.646**	**0.642**
Ours (pairwise MRF)	0.833	0.830	0.700	0.696	0.633	0.629
Ours (3×3 FOE)	**0.850**	**0.846**	**0.711**	0.707	0.640	0.637

Table 4.1: Average deblurring results for 64 test images. *GT* denotes that the ground truth noise parameter σ was used, *NE* indicates that σ was assumed to be unknown and an estimate was used instead. For all methods except ours, the approach of Zoran and Weiss (2009) was used for noise estimation prior to deblurring.

to one of 256 discrete intensities. We deblurred with a slightly perturbed version of the true blur kernel by adding white Gaussian noise with variance 10^{-6} to it. The motivation is to mimic a more realistic scenario where the estimated blur kernel always contains some error.

We compare our approach against the standard MAP-based methods of Levin et al. (2007) and Krishnan and Fergus (2009), as well as the classical Lucy-Richardson algorithm (Richardson, 1972; Lucy, 1974). Both Levin et al. (2007) and Krishnan and Fergus (2009) use image priors with hyper-Laplacian potential functions $e^{-\rho(x)}$ with $\rho(x) = |x|^{\alpha}$. We used the original implementation of Levin et al. (2007) unchanged with $\alpha = 4/5$, and chose $\alpha = 2/3$ for the original implementation of Krishnan and Fergus (2009) to achieve best performance. In addition, we also applied the 5×5 FOE prior from Roth and Black



(a) $\sigma = 2.55$ (b) $\sigma = 7.65$

(c) $\sigma = 12.75$

Figure 4.2: Sorted PSNR differences for the 64 deblurred images between our method (3×3 FOE) and all others. Noise estimates (using the method of Zoran and Weiss (2009)) instead of the ground truth σ were used for the other approaches. Everything below the black dashed line means that our method (3×3 FOE) was better than the competing one. *Best viewed in color.*

(2009) to the non-blind deblurring case, and used an expectation maximization (EM) algorithm (see Levi, 2009) to perform MAP estimation.

For all competing methods, the regularization parameters were determined (per noise level) on a separate training set of 16 images, using each of the eight blur kernels from Levin et al. (2009) twice. Since our approach does not necessarily require the noise level as input, we also instantiated competing methods with noise estimates, which are here acquired by the algorithm of Zoran and Weiss (2009).

The quantitative deblurring results for the 64 images and three noise levels are summarized in Table 4.1. Figure 4.2 visually illustrates the performance difference of our approach to various baselines. The results show that Lucy-Richardson (Richardson, 1972; Lucy, 1974) and the very fast method of Krishnan and Fergus (2009) are far behind the other methods for all noise levels, particularly behind ours (at least 0.5dB worse on average). We find that our method particularly outperforms competing approaches for the smallest – and for many applications likely the most realistic – noise level. For example, the widely used method of Levin et al. (2007) is outperformed by about 0.6dB. An interesting observation is also that deblurring with the 5×5 FOE prior from Roth and Black (2009) outperforms Levin et al. (2007). This shows that the learned, high-order MRF prior is at least partly responsible for the observed performance difference

to Levin et al. (2007). Nonetheless, our Bayesian approach with integrated noise estimation still performs about 0.25dB better than using a 5×5 FOE with MAP estimation, despite using smaller 3×3 cliques. Interestingly, even when combined with a simple pairwise MRF our method comes close in performance to the 5×5 FOE. This demonstrates that the proposed MMSE approach improves performance.

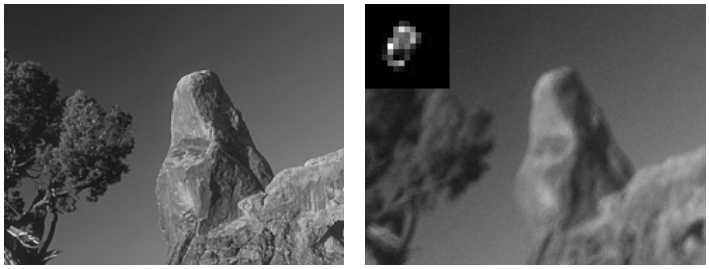
A qualitative comparison on two larger images that were not part of the test set can be seen in Figures 4.3 and 4.4. We observe in both images that our method is especially good at preserving textural detail, while at the same time allowing for very smooth regions, which appear too crisp or show artifacts in case of the competing methods (*e.g.* the sky in Figure 4.3 or the background near the top of Figure 4.4).

COMPUTATIONAL CONSIDERATIONS. It is not surprising that the Bayesian approach taken in this chapter is computationally more expensive than MAP estimation (energy minimization). The reported results are intended to show the performance limits of our method and consequently use conservative thresholds for assessing sampler convergence. For practical purposes it is easily possible to relax the convergence criteria with little to no influence on the deblurring performance. For the 64 test images in Table 4.1 and $\sigma = 2.55/7.65/12.75$, a simple MATLAB implementation achieves an average runtime of 4.6/3.9/3.9 minutes, while staying within 0.03dB of the quoted results. A similar MAP implementation used for the 5×5 FOE (Roth and Black, 2009) is only about 4 times faster. The methods of Krishnan and Fergus (2009) and Levin et al. (2007) are optimized for speed and need at most a few seconds per image. While much faster, their deblurring performance is significantly worse (1.7dB and 0.6dB). Furthermore, they require additional effort for determining suitable regularization parameters.

4.7 CONCLUSIONS AND FUTURE WORK

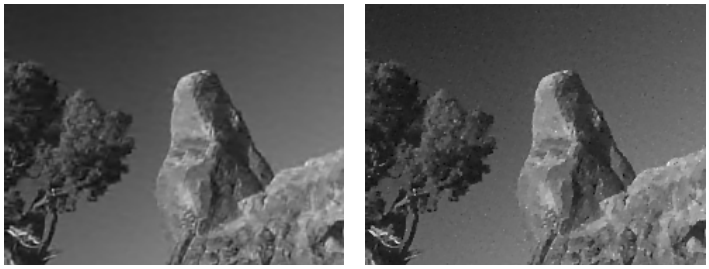
Based on posterior sampling, we presented a comprehensive Bayesian framework for non-blind deblurring. We achieved improved application performance by exploiting learned, high-order priors and relying on MMSE estimation. In an experimental evaluation, our framework was demonstrated to outperform conventional MAP deblurring algorithms using standard quantitative measures. A further merit of the proposed approach is the integrated treatment of unknown parameters such as the noise standard deviation.

Future work in Bayesian inference needs to engineer less computationally intensive sampling algorithms. Another avenue for future research is to estimate additional parameters of the problem at hand; for deblurring in particular, estimating a non-parametric blur kernel in an integrated fashion would extend our framework to the problem



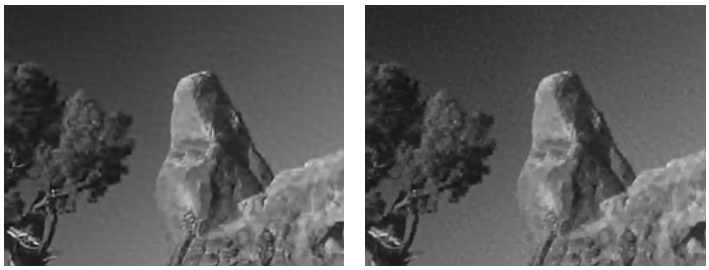
(a) Original

(b) Blurred input image, PSNR = 23.89dB, SSIM = 0.588



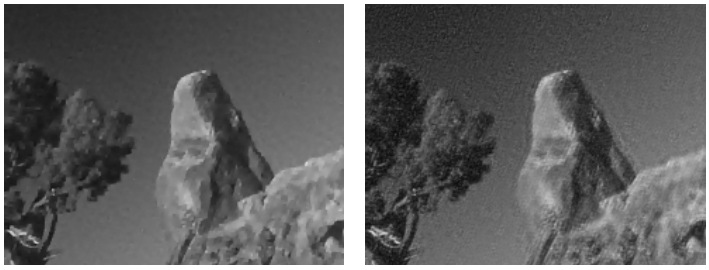
(c) Ours (3×3 FOE), PSNR = 29.05dB, SSIM = 0.864

(d) Ours (pairw. MRF), PSNR = 28.84dB, SSIM = 0.842



(e) 5×5 FOE (MAP), model of Roth and Black (2009), PSNR = 28.81dB, SSIM = 0.844

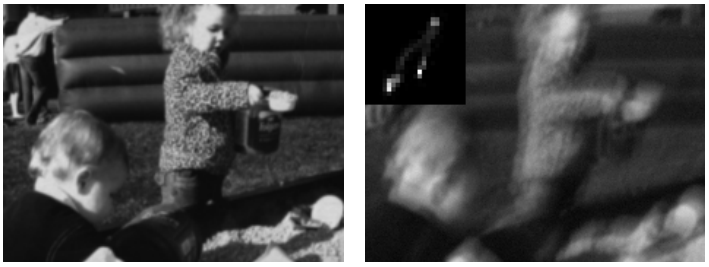
(f) Levin et al. (2007), PSNR = 28.54dB, SSIM = 0.826



(g) Krishnan and Fergus (2009), PSNR = 28.38dB, while SSIM = 0.825

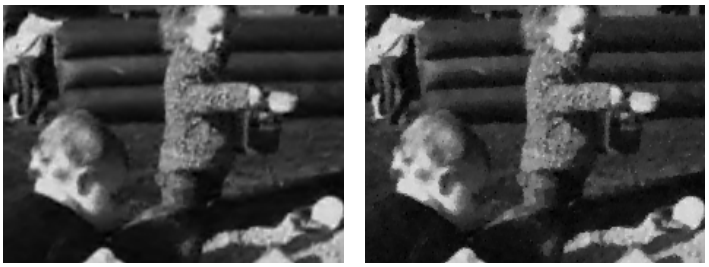
(h) Method of Richardson (1972) and Lucy (1974), PSNR = 27.01dB, SSIM = 0.693

Figure 4.3: Deblurring example (cropped; image from BSDS). Our method simultaneously preserves rock texture and smooth sky. The 15×15 blur kernel shown in *(b)* is spatially resized and its entries are scaled for better visualization. *Best viewed on screen.*



(a) Original

(b) Blurred input image, PSNR = 19.24dB, SSIM = 0.480

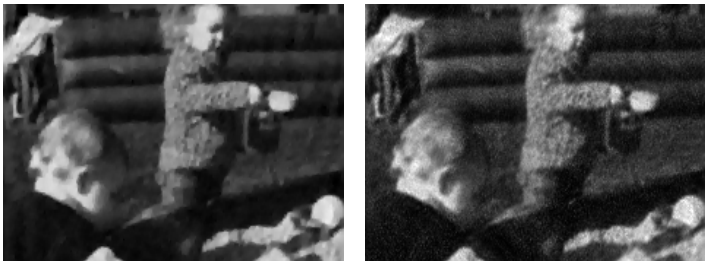


(c) Ours (3×3 FOE), PSNR = 32.09dB, SSIM = 0.920

(d) Ours (pairwise MRF), PSNR = 30.31dB, SSIM = 0.878

(e) 5×5 FOE (MAP), model of Roth and Black (2009), PSNR = 31.71dB, SSIM = 0.914

(f) Levin et al. (2007), PSNR = 31.33dB, SSIM = 0.898



(g) Krishnan and Fergus (2009), PSNR = 28.30dB, while SSIM = 0.873

(h) Method of Lucy (1974) and Richardson (1972), PSNR = 26.23dB, SSIM = 0.708

Figure 4.4: Deblurring example (cropped; image from Levin et al. (2009)). The blur exhibits sparseness typical to camera shake. Our method simultaneously preserves clothing texture and smooth background regions. The 23×23 blur kernel shown in *(b)* is spatially resized and its entries are scaled for better visualization. *Best viewed on screen.*

of blind deblurring. Finally, future work should also be devoted to gathering ground-truth data for spatially varying blur and evaluating our and other approaches on this task.

5

MEAN FIELD FOR CONTINUOUS HIGH-ORDER MRFS

CONTENTS

Probabilistic inference beyond maximum a-posteriori (MAP) estimation is of interest in computer vision, both for learning appropriate models and in applications. Yet, common approximate inference techniques, such as belief propagation (BP), have largely been limited to discrete-valued Markov random fields (MRFs) and models with small cliques. Oftentimes, neither is desirable from an application standpoint. This chapter studies mean field inference for continuous-valued MRF models with high-order cliques. Mean field can be applied effectively to such models by exploiting that the factors of certain classes of MRFs can be formulated using Gaussian mixtures, which allows retaining the mixture indicator as a latent variable. We use an image restoration setting to show that resulting mean field updates have a computational complexity quadratic in the clique size, which makes them scale even to large cliques. We contribute an empirical study with four applications: Image denoising, non-blind deblurring, noise estimation, and layer separation from a single image. We find mean field to yield a favorable combination of performance and efficiency, *e.g.* outperforming MAP estimation in denoising while being competitive with expensive sampling approaches. Novel approaches to noise estimation and layer separation demonstrate the breadth of applicability.

5.1 INTRODUCTION

Probabilistic models have found widespread use in all areas of computer vision. Applying them to a concrete problem entails computing a solution by means of inference. The approach used most widely is MAP estimation, which corresponds to a Bayes-optimal prediction under the 0/1-loss (see *e.g.* Pletscher et al. (2011)). In many applications the 0/1-loss is not appropriate, since it penalizes all incorrect solu-

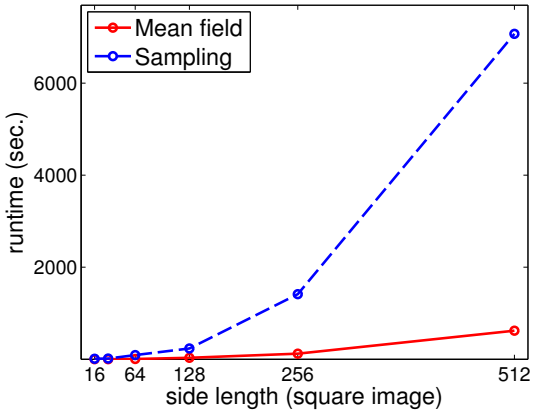


Figure 5.1: Empirical runtime comparison of mean field and Gibbs sampling (Schmidt et al., 2010) for denoising with an MRF with 3×3 cliques.

tions equally. One alternative is to compute the Bayesian minimum mean squared error (MMSE) estimate, which amounts to estimating the posterior mean (Portilla et al., 2003). However, this is challenging in the context of MRF models[1]. Standard approaches to computing the posterior mean are based on sum-product BP (Lasowski et al., 2011; Potetz, 2007) or sampling algorithms (Schmidt et al., 2010), which often come at a high computational cost.

In this chapter we study mean field (Geiger and Girosi, 1991) as efficient alternative for MRFs with continuous variables. To that end we follow Levin et al. (2011) to assume that the factors can be represented as Gaussian mixture models, which allows retaining the mixture indices as explicit latent variables. These in turn make it convenient to derive simple update equations. But unlike Levin et al. (2011), our focus is on MRFs of high-order, which have found increased adoption in recent years (Ishikawa, 2009; Roth and Black, 2009). In particular, we argue and show that for a certain class of problems, one of the key advantages of the mean field method is its computational efficiency even in the presence of high-order factors: The effort of an update cycle scales quadratically in the clique size of the MRF. This is in contrast to the runtime of classical, discrete BP, which scales exponentially in the clique size (Potetz, 2007).

To demonstrate the practical benefits of using mean field for inference in continuous high-order MRFs, we study four applications in the context of image restoration, where accurate probabilistic models are available (e.g. Schmidt et al. (2010)). Using image denoising and non-blind deblurring as a testbed, we demonstrate that mean field inference yields a favorable combination of efficiency and perfor-

1 Even though we focus on MRFs, our discussion also applies to conditional random fields.

mance, outperforming MAP estimation while being much faster than sampling-based inference. We also contribute novel mean field methods for noise estimation and layer separation. Noise estimates are obtained by maximizing a marginalized density using the expectation maximization (EM) algorithm and mean field to approximate intractable expectations. Layer separation from a single image is a massively inverse problem with a multimodal likelihood. One novelty of our mean field approach is the extraction of an explicit noise layer.

5.2 RELATED WORK

Geiger and Girosi (1991) were among the first to apply mean field to MRFs in vision. In particular, they used formulations based on a line process. Our approach is related in that it also uses discrete latent variables, but employs multinomial rather than binary ones, which correspond to the component indices of a Gaussian mixture. Moreover, the work of Geiger and Girosi (1991) is limited to pairwise MRFs, whereas we apply mean field to high-order MRFs.

Although recent work puts forward efficient algorithms for discrete mean field (Krähenbühl and Koltun, 2011), this chapter focuses on continuous variables, combined with learned mixture model priors. Mean field has been used for continuous MRFs with mixture model potentials, for example in deblurring (Chantas et al., 2008; Levin et al., 2011), reflectometry (Romeiro and Zickler, 2010), and layer separation of cartoon images (Miskin and MacKay, 2000). In contrast to previous work, we study the impact of mean field for MMSE estimation in generatively trained, high-order models, derive a polynomial update complexity for a certain class of problems, and develop new methods for noise estimation and natural image layer separation. We also provide a quantitative evaluation of mean field for non-blind deblurring, whereas previous work in deblurring has focused on mean field for kernel estimation in the gradient domain (Fergus et al., 2006; Levin et al., 2011) instead of image restoration in the spatial domain; Section 5.4 demonstrates the latter.

One advantage of continuous variables and mixture models is that the mean field update scheme scales quadratically in the clique size for typical problems of image restoration. This is notable compared to other approximate inference algorithms that allow moment estimation, such as BP. In the classical, discrete case, the factor-to-variable messages of BP involve a brute-force summation, which incurs an exponential update complexity in the clique size (Potetz, 2007). While BP can be sped up for certain model classes, such techniques have been limited to pairwise potentials (Felzenszwalb and Huttenlocher, 2006; Lasowski et al., 2011) or graphs with 2×2 cliques (Lan et al., 2006; Potetz, 2007) in practice. Non-parametric BP (Sudderth et al., 2003) remains challenging due to the need to approximate a product

of Gaussian mixtures of exponential size for every message computation. On the other hand, Gibbs sampling for mixture model priors (Schmidt et al., 2010) scales poorly to larger inputs (Figure 5.1), and determining convergence of Monte Carlo methods is not always straightforward. This stands in contrast to mean field, which is guaranteed to converge (Minka, 2005).

5.3 MEAN FIELD FOR CONTINUOUS HIGH-ORDER MRFS

FREE ENERGY. A key challenge in probabilistic inference is to compute moments or modes of probabilities, whose exact computation is often prohibitive, necessitating approximate inference techniques. The objective of mean field is to approximate an intractable model by a tractable density. The moments of the approximate density then serve as approximations for the true moments. Given a true, intractable posterior $p(\mathbf{x}|\mathbf{y})$ the approximation $q(\mathbf{x})$ is determined by minimizing the Kullback-Leibler divergence $\mathrm{KL}(q(\mathbf{x})\|p(\mathbf{x}|\mathbf{y}))$ (Minka, 2005). This can be achieved by minimizing the free energy

$$F(q) = -\int q(\mathbf{x}) \log p(\mathbf{x}, \mathbf{y}) d\mathbf{x} + \int q(\mathbf{x}) \log q(\mathbf{x}) d\mathbf{x}, \tag{5.1}$$

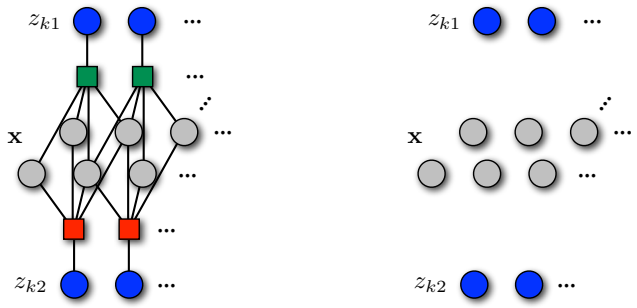
as follows from Equations (2) and (3) of Winn and Bishop (2005).

MRFS WITH MIXTURE POTENTIALS. Even though mean field is in principle applicable to arbitrary continuous MRFs where the factors are modeled as Gaussian mixtures, we choose a specific model family for concreteness. We use a variant of Field of Experts (FOE) (Schmidt et al., 2010; Weiss and Freeman, 2007), a continuous high-order MRF in which the potentials are formulated as a Gaussian scale mixture (GSM) (Wainwright and Simoncelli, 2000). This not only provides us with a suitable testbed for mean field inference, but after expansion with indicator variables also fulfills the requirement that the potentials are in the exponential family (Minka, 2005). This class of image priors is based on learned, zero-mean, linear image filters $\mathbf{F}_\gamma$, whose response at factor (clique) $\mathbf{x}_k$ of the image $\mathbf{x}$ is modeled by a learned Gaussian scale mixture expert (potential)

$$\phi_\gamma(\mathbf{F}_\gamma^T \mathbf{x}_k) = \sum_{j=1}^{J} \pi_{\gamma j} \mathcal{N}\left(\mathbf{F}_\gamma^T \mathbf{x}_k | 0, \sigma_{\gamma j}^2\right). \tag{5.2}$$

The filters $\mathbf{F}_\gamma$ have zero mean in order to be invariant to global gray-level shifts. Multiplying over all filters and cliques, and a broad Gaussian factor ($\epsilon = 10^{-8}$) yields the prior

$$p(\mathbf{x}) \propto e^{-\epsilon \|\mathbf{x}\|^2} \prod_k \prod_\gamma \phi_\gamma(\mathbf{F}_\gamma^T \mathbf{x}_k). \tag{5.3}$$



(a) Factor graph of Equation (5.4)

(b) Fully-factorized graph

Figure 5.2: Illustration of the high-order factor graph representation (a) and the fully factorized approximation (b) for a MRF with 2×2 cliques and 2 filters.

Following Levin et al. (2011), GSMs (and any Gaussian mixture) can be made tractable for mean field inference by expanding with indicator variables, which represent the indices of the mixture components. Each pair of clique and filter (expert) is equipped with a discrete variable $z_{k\gamma} \in \{1, ..., J\}$ such that $p(z_{k\gamma}) = \pi_{\gamma z_{k\gamma}}$. The resulting, fully expanded prior is

$$p(\mathbf{x}, \mathbf{z}) \propto e^{-\epsilon \|\mathbf{x}\|^2} \prod_k \prod_\gamma \pi_{\gamma z_{k\gamma}} \mathcal{N}\left(\mathbf{F}_\gamma^T \mathbf{x}_k | 0, \sigma^2_{\gamma z_{k\gamma}}\right), \tag{5.4}$$

with the property that $\sum_{\mathbf{z}} p(\mathbf{x}, \mathbf{z}) = p(\mathbf{x})$. Figure 5.2(a) shows an example factor graph for this model. Note that after augmenting with indicator varibles the model continues to be of high order, since the indicators $z_{k\gamma}$ are shared across the whole clique, preventing further factorization. The key benefit is that the prior of Equation (5.4) is in the exponential family and thus readily admits mean field inference.

MEAN FIELD APPROXIMATION. As concrete application we consider Bayesian image restoration $p(\mathbf{x}|\mathbf{y}, \theta) \propto p(\mathbf{y}|\mathbf{x}, \theta)p(\mathbf{x})$ of an observed image $\mathbf{y}$, where $p(\mathbf{y}|\mathbf{x}, \theta)$ is a likelihood model with parameters θ. For example, θ may be a blur kernel $\mathbf{K}$ in image deblurring. We make the assumption that the likelihood $p(\mathbf{y}|\mathbf{x}, \theta)$ is proportional to a Gaussian in $\mathbf{x}$ with precision matrix $\mathbf{P}$ and mean $\mathbf{P}^{-1}\mathbf{b}$ (see Section 5.4 for concrete examples). Augmented by indicator variables, the Bayesian formulation takes the form $p(\mathbf{x}, \mathbf{z}|\mathbf{y}, \theta) \propto p(\mathbf{y}|\mathbf{x}, \theta)p(\mathbf{x}, \mathbf{z})$. The conventional choice of mean field approximation is the fully-factorized model

$$q(\mathbf{x}, \mathbf{z}) = q(\mathbf{x}) \prod_k \prod_\gamma q(z_{k\gamma}), \tag{5.5}$$

where $q(\mathbf{x}) = \mathcal{N}(\boldsymbol{\mu}, \mathbf{C})$ with a diagonal covariance matrix $\mathbf{C}$. Figure 5.2(b) shows the corresponding factor graph. We here follow the

fully-factorized approach. Mean field proceeds by updating variables in turn, while holding the others fixed. Detailed update equations arise by inserting $q(\mathbf{x}, \mathbf{z})$ into the free energy from Equation (5.1), integrating and differentiating. The updates resemble the case of pairwise MRFs (see Levin et al. (2011) for detailed derivations), with the difference being that the filters $\mathbf{F}_\gamma$ cover more than two pixels. The update of the indicator distributions $q(z_{k\gamma})$ has linear complexity in the clique size and number of pixels:

$$q(z_{k\gamma} = j_0) \propto \frac{\pi_{\gamma j_0}}{\sigma_{\gamma j_0}} \exp\left\{-\frac{\mathrm{E}[(\mathbf{F}_\gamma^\mathsf{T}\mathbf{x}_k)^2]}{2\sigma^2_{\gamma j_0}}\right\}. \tag{5.6}$$

On the other hand, to update $q(\mathbf{x}) = \mathcal{N}(\boldsymbol{\mu}, \mathbf{C})$ we require the matrix

$$\mathbf{A} = \mathbf{P} + \epsilon\mathbf{I} + \sum_\gamma \mathbf{T}^\mathsf{T}_{\mathbf{F}_\gamma} \mathbf{W}_\gamma \mathbf{T}_{\mathbf{F}_\gamma}. \tag{5.7}$$

The ensuing text sheds light on how the matrix $\mathbf{A}$ is used to update the pixel distributions. Here, the $\mathbf{W}_\gamma$ denote diagonal matrices with positive entries (Levin et al., 2011) whereas the $\mathbf{T}_{\mathbf{F}_\gamma}$ denote convolution matrices, such that $\mathbf{T}_{\mathbf{F}_\gamma}\mathbf{x} \equiv \mathbf{F}_\gamma \otimes \mathbf{x}$. The matrix $\mathbf{A}$ is symmetric and positive definite. Updating the diagonal covariance $\mathbf{C}$ of the approximate $q(\mathbf{x}) = \mathcal{N}(\boldsymbol{\mu}, \mathbf{C})$ then proceeds by inverting the diagonal elements of $\mathbf{A}$, such that $C_{ii} = (A_{ii})^{-1}$. Since computing the diagonal elements of $\mathbf{A}$ needs only linear cost in the clique size and number of pixels, this step is also efficient. The bottleneck of each update cycle lies in updating the mean $\boldsymbol{\mu}$, which requires minimizing the quadratic objective

$$Q(\boldsymbol{\mu}) = \frac{1}{2}\boldsymbol{\mu}^\mathsf{T}\mathbf{A}\boldsymbol{\mu} - \mathbf{b}^\mathsf{T}\boldsymbol{\mu}. \tag{5.8}$$

We here quantify this effort as follows.

POLYNOMIAL UPDATES. The mean field algorithm updates each component of $\boldsymbol{\mu}$ sequentially while holding the others fixed. This update scheme for the elements of $\boldsymbol{\mu}$ corresponds to a Gauss-Seidel solver of the system $\mathbf{A}\boldsymbol{\mu} = \mathbf{b}$ (Levin et al., 2011). It is important to note that the clique size $\mathcal{C}$ of the MRF $p(\mathbf{x}|\mathbf{y}, \boldsymbol{\theta})$ is given by the size of the largest prior filter among $\{\mathbf{F}_\gamma\}$, whereas the Gaussian likelihood contributes $\mathcal{P}$ pairwise connections to other pixels. The update of μ_i costs $\mathcal{O}(\mathcal{C}^2 + \mathcal{P})$ operations (equating the derivative $\frac{\mathrm{d}}{\mathrm{d}\mu_i}Q(\boldsymbol{\mu})$ to zero and solving yields the update equation). By contrast, an update of classical, discrete BP in an MRF has *exponential* complexity in the clique size (Potetz, 2007), which is prohibitively slow for high-order cliques and large label spaces.

5.4 APPLICATIONS AND EXPERIMENTS

To study the application of mean field to continuous high-order MRFs we consider four different applications in image restoration. We use

Fields of Experts as image priors, which have served as a testbed for various studies on approximate inference methods (Ishikawa, 2009; Lan et al., 2006; Potetz, 2007). Specifically, we use the 3×3 FOE and pairwise MRF of Schmidt et al. (2010), which allow to assess efficiency and runtime of inference methods for high-order models.

IMAGE RESTORATION. We consider the following generic formulation for image restoration: Given a blur kernel $\mathbf{K}$ and degraded image $\mathbf{y}$, recover $\mathbf{x}$ such that $\mathbf{y} = \mathbf{K} \otimes \mathbf{x} + \mathbf{n}$. Here, we make the widespread assumption of additive Gaussian noise $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$. It is convenient to represent the blur using a convolution matrix $\mathbf{T}_\mathbf{K}$. Denoising is a special case of $\mathbf{T}_\mathbf{K} = \mathbf{I}$ being the identity matrix. A Bayesian approach gives rise to $p(\mathbf{x}|\mathbf{y}, \mathbf{T}_\mathbf{K}) \propto p(\mathbf{y}|\mathbf{x}, \mathbf{T}_\mathbf{K}) p(\mathbf{x})$, where the likelihood takes the form $p(\mathbf{y}|\mathbf{x}, \mathbf{T}_\mathbf{K}) = \mathcal{N}(\mathbf{y}|\mathbf{T}_\mathbf{K}\mathbf{x}, \sigma^2 \mathbf{I})$. For non-trivial blur ($\mathbf{T}_\mathbf{K} \neq \mathbf{I}$), the posterior is highly connected – hence efficient inference is paramount. For image restoration, Equation (5.7) takes the form

$$\mathbf{A} = \sigma^{-2} \mathbf{T}_\mathbf{K}^\top \mathbf{T}_\mathbf{K} + \epsilon \mathbf{I} + \sum_\gamma \mathbf{T}_{\mathbf{F}_\gamma}^\top \mathbf{W}_\gamma \mathbf{T}_{\mathbf{F}_\gamma}. \tag{5.9}$$

To evaluate the applicability of continuous mean field to denoising, we compare to MAP estimation on the one hand, and on the other to MMSE estimation using Gibbs sampling based on the models from Schmidt et al. (2010). The motivation is twofold: Comparing with MAP estimation in generative models establishes a baseline to be outperformed by MMSE estimation via approximate inference[2]. A comparison with Gibbs sampling allows to assess the potential advantages in terms of efficiency (see Figure 5.1). Additionally, we also compare with MAP estimation in a standard pairwise MRF with Laplacian potentials: $\phi(y) = \exp(-(y^2 + \epsilon)^{\frac{1}{2}})$ for small ϵ. These potentials are in wide use in vision (e.g. Levin and Weiss (2007)) and are also closely related to popular total variation (TV) regularizers (Rudin et al., 1992). This puts the performance of mean field into relation with a model/inference combination standard to vision. For both MAP and mean field, we use a regularization parameter $\lambda \in (0, 1)$, such that the influences of prior and likelihood are calibrated[3] via $p(\mathbf{y}|\mathbf{x}, \mathbf{T}_\mathbf{K})^\lambda p(\mathbf{x})^{1-\lambda}$. Notably, the parameter λ for mean field is trained on a separate set of images, whereas for the MAP methods, it is *optimized on the test set* to demonstrate the highest achievable performance.

The results are summarized in Table 5.1. Mean field not only outperforms the baseline of MAP estimation in the generative model,

2 We note that it is possible to incorporate the loss function during training of a random field (Pletscher et al., 2011), but our focus is on Bayesian prediction with probabilistic models.

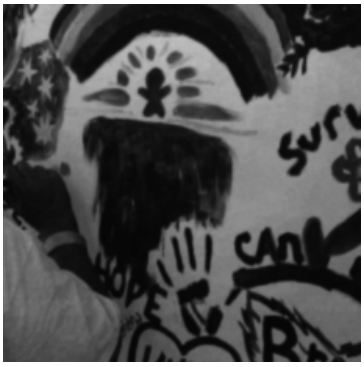
3 Although the use of a regularization parameter changes the model and weakens the principled Bayesian approach, it is likely indispensable to mitigate the effects of minimizing exclusive divergence (Minka, 2005).

Table 5.1: Denoising results on 10 BSDS images reduced to half of their initial size (Lan et al., 2006; Schmidt et al., 2010). Average PSNR (dB) and runtime for different noise levels. The term "w/ λ" denotes usage of a regularization parameter, while "Lapl." abbreviates the Laplacian model. Average runtime measured on a 2.67GHz Core i7 processor.

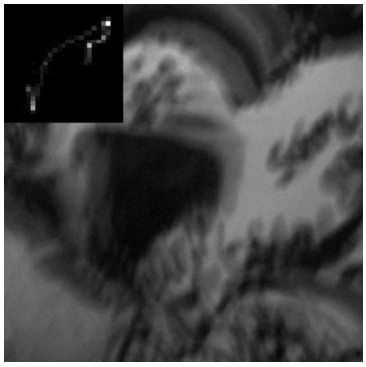
		$\sigma = 5$		$\sigma = 10$		$\sigma = 20$	
		PSNR	t	PSNR	t	PSNR	t
pairw. MRF	CG w/ λ	35.61	8m	30.84	9m	26.55	9m
	Mean field w/ λ	**36.65**	**8s**	**32.22**	44s	28.26	**22s**
	Gibbs sampling	36.41	57s	32.09	2m	**28.32**	3m
pairw. Lapl.	CG w/ λ	36.36	**8s**	31.91	**12s**	28.11	23s
3×3 FOE	CG w/ λ	36.83	9m	32.19	16m	27.98	28m
	Mean field w/ λ	36.83	**89s**	32.36	**63s**	28.40	**72s**
	Gibbs sampling	**37.23**	8m	**32.85**	12m	**28.91**	18m

but also does better than the (best-case) Laplacian MRF. To compare with Gibbs sampling, we use the publicly available code of Schmidt et al. (2010). In the pairwise MRF, mean field is competitive with Gibbs sampling, sometimes even outperforming it, while being many times faster. It is interesting to note that the regularization parameter λ helps overcoming some of the deficiencies of the mean field approximation. In case of a 3×3 model, mean field falls somewhat below the performance of Gibbs sampling, most likely due to the full factorization of the approximate distribution being less accurate here. Nonetheless, we find a clear improvement over the pairwise model without incurring a large penalty in terms of computational efficiency; mean field remains many times faster than Gibbs sampling. We also remark that mean field inference even for the *pairwise* MRF outperforms the results obtained for a *high-order* 2×2 FOE model using graph cuts (Ishikawa, 2009) (8–12 minutes on a 2.33GHz Xeon E5345 processor), as well as sum-product (Potetz, 2007) (30–60 minutes on a 2.2GHz Opteron 275) and max/sum-product BP (Lan et al., 2006) (8 hours on a 3GHz Xeon) in terms of image quality and efficiency. Note that the runtime is also competitive with MAP estimation by conjugate gradient (CG) descent in the simple pairwise Laplacian model, although specifically tailored optimization techniques may improve the runtime of both methods. As further runtime comparison to Gibbs sampling, we test both methods on a sequence of square images of ascending size $(2^4, 2^5, 2^6 ..., 2^9)$. Figure 5.1 demonstrates empirically that mean field scales well to large images, whereas Gibbs sampling quickly incurs high runtime.

With regard to non-blind image deblurring, we use a benchmark of 64 images, each blurred with one of 8 realistic blur kernels (Levin



(a) Ground truth image

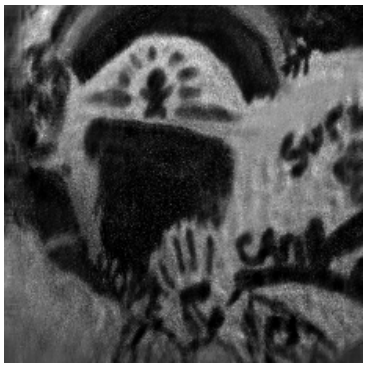


(b) Blurred, 1% noise level, PSNR 17.15, SSIM .339

(c) Mean field, PSNR 29.53, SSIM .876

(d) Levin et al. (2007), PSNR 29.33, SSIM .841

(e) Krishnan and Fergus (2009), PSNR 26.47, SSIM .822



(f) Lucy (1974), PSNR 24.20, SSIM .621

Figure 5.3: Deblurring on a standard image and blur kernel (Levin et al., 2011) with 1% additive noise ($\sigma = 2.55$). The 27×27 blur kernel is displayed in the upper left corner of the blurry image, enlarged and scaled to full intensity. The result of mean field has sharp details and fewer ringing artifacts.

Table 5.2: Deblurring results on 64 test images and five noise levels corresponding to $1\% - 5\%$ of the highest intensity, 255. PSNR (dB), and SSIM values denote averages over all images. Mean field performs better on the majority of noise levels.

Performance in PSNR

Method	$\sigma = 1\%$	$\sigma = 2\%$	$\sigma = 3\%$	$\sigma = 4\%$	$\sigma = 5\%$
Lucy (1974)	25.38	23.27	21.85	20.84	19.83
Krishnan and Fergus (2009)	26.97	25.69	24.91	24.34	23.93
Levin et al. (2007)	28.03	26.28	25.36	24.74	**24.29**
Mean field	**28.20**	**26.42**	**25.44**	**24.77**	24.25

Performance in SSIM

Method	$\sigma = 1\%$	$\sigma = 2\%$	$\sigma = 3\%$	$\sigma = 4\%$	$\sigma = 5\%$
Lucy (1974)	.703	.542	.423	.334	.224
Krishnan and Fergus (2009)	.800	.724	.671	.632	.608
Levin et al. (2007)	.823	.741	.689	**.652**	**.625**
Mean field	**.833**	**.747**	**.692**	.650	.616

et al., 2011). We consider additive noise levels σ corresponding to a range of $1\% - 5\%$ of the maximum gray intensity 255. Table 5.2 summarizes quantitative results over 64 images; MMSE estimation with mean field outperforms standard non-blind MAP methods (Krishnan and Fergus, 2009; Levin et al., 2007) on the majority of noise levels. Moreover, Figure 5.3 illustrates a qualitative advantage: Mean field yields a deblurring result with both sharp details and fewer ringing artifacts. Previous work in deblurring (*e.g.*, Fergus et al. (2006)) has largely focused on using mean field for kernel estimation in the gradient domain instead of image restoration in the spatial domain, as we demonstrate here.

NOISE ESTIMATION. Given an input image $\mathbf{y} = \mathbf{x} + \mathbf{n}$, the task in noise estimation is to determine the standard deviation of (assumed) Gaussian noise $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$ on the latent image $\mathbf{x}$. Much like Levin et al. (2011) for kernel estimation, we use the EM algorithm to maximize the incomplete-data likelihood $p(\mathbf{y}|\sigma)$, which under the assumption of a flat prior $p(\sigma) \propto 1$ also maximizes $p(\sigma|\mathbf{y})$. It is convenient to choose the factorization $p(\mathbf{x}, \mathbf{y}, \sigma) = p(\mathbf{y}|\mathbf{x}, \sigma)p(\mathbf{x})$, using the standard likelihood $p(\mathbf{y}|\mathbf{x}, \sigma) = \mathcal{N}(\mathbf{x}, \sigma^2\mathbf{I})$. The E–step requires evaluating the moments of $p(\mathbf{x}|\mathbf{y}, \sigma_{\mathtt{old}}) \propto p(\mathbf{y}|\mathbf{x}, \sigma_{\mathtt{old}})p(\mathbf{x})$, which is intractable for loopy MRFs. Instead, we leverage mean field to compute an approxi-

Table 5.3: Noise estimation results on 68 test images (Roth and Black, 2009). Relative absolute error (Zoran and Weiss, 2009) (RAE $= |\sigma_{est} - \sigma|/\sigma$), and relative squared error (RSE $= |\sigma_{est} - \sigma|^2/\sigma^2$) are reported in percent. The entries denote averages over all images.

Performance in RAE

Method	$\sigma = 10$	$\sigma = 20$	$\sigma = 30$	$\sigma = 40$	$\sigma = 50$	$\sigma = 60$
Zoran and Weiss (2009)	**12.58**	9.04	8.98	10.31	12.46	14.96
EM & mean field	14.34	**9.00**	**6.53**	**5.96**	**7.24**	**9.66**

Performance in RSE

Method	$\sigma = 10$	$\sigma = 20$	$\sigma = 30$	$\sigma = 40$	$\sigma = 50$	$\sigma = 60$
Zoran and Weiss (2009)	**2.41**	**1.17**	1.20	1.51	2.03	2.69
EM & mean field	4.76	1.95	**0.96**	**0.70**	**0.83**	**1.25**

mation $q(\mathbf{x})$ of $p(\mathbf{x}|\mathbf{y}, \sigma_{old})$. The M–step consists in updating the noise estimate by

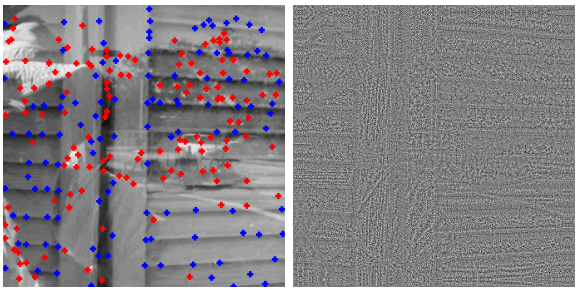
$$\sigma_{new} = \operatorname*{argmax}_{\sigma} \mathrm{E}_q[\log p(\mathbf{x}, \mathbf{y}|\sigma)] = \sqrt{\frac{\mathrm{E}_q[\|\mathbf{x}-\mathbf{y}\|^2]}{N}}, \tag{5.10}$$

where N denotes the number of elements in $\mathbf{x}$ and $\mathrm{E}_q[\|\mathbf{x}-\mathbf{y}\|^2] = \sum_{i=1}^{N} \mu_i^2 + C_{ii} - 2\mu_i y_i + y_i^2$, for the approximate probability $q(\mathbf{x}) = \mathcal{N}(\boldsymbol{\mu}, \mathbf{C})$. We test on a benchmark of 68 images and find superior results on multiple noise levels (Table 5.3) compared to a recent, state of the art method (Zoran and Weiss, 2009), which leverages a connection between white noise and kurtosis values. The results of Table 5.3 are obtained for the high-order 3×3 FOE, and although many EM iterations are required, mean field admits efficient inference.

LAYER SEPARATION. To show the versatility of mean field as inference algorithm, we lastly consider the problem of recovering two hidden images given their sum $\mathbf{y} = \mathbf{x}_1 + \mathbf{x}_2$, which can occur when light is reflected off glass. This massively inverse problem may be alleviated by user input (Levin and Weiss, 2007). Here, the user marks a set of points S_1 and S_2 of the input image $\mathbf{y}$ as edges belonging to layers $\mathbf{x}_1$ and $\mathbf{x}_2$, respectively: Figure 5.4 depicts a problem instance. We denote by $\mathbf{S}_1$ and $\mathbf{S}_2$ the diagonal matrices that pick out image elements corresponding to the indices collected in S_1 and S_2, such that *e.g.*, $\|\mathbf{S}_1\mathbf{y}\|^2 = \sum_{k \in S_1} y_k^2$. These matrices allow to encode the user marks into Gaussians: For zero-mean prior filters $\mathbf{F}_\gamma$, we model the user input by

$$p(\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2, S_1, S_2) = \mathcal{N}(\mathbf{x}_1 + \mathbf{x}_2, \sigma^2\mathbf{I}) \prod_{k=1,2} \mathcal{N}(\mathbf{x}_k, \mathbf{M}_k^{-1}), \tag{5.11}$$

(a) Input image (b) First inferred layer (c) Second inferred layer



(d) User marks (e) Noise layer

Figure 5.4: Layer separation by mean field inference of a standard input mixture (Levin and Weiss, 2007). The input image is corrupted by additive Gaussian noise of standard deviation 1. Our method recovers two hidden layers and pushes noise into a residual layer by mean field inference. The layers are scaled to the full intensity range for better visualization. *Best viewed on screen.*

where the matrices $\mathbf{M}_k = \sum_\gamma \left(\mathbf{S}_k \mathbf{T}_{\mathbf{F}_\gamma}\right)^T \left(\mathbf{S}_k \mathbf{T}_{\mathbf{F}_\gamma}\right)$ constrain the filter responses of layer $\mathbf{x}_k$ to be close to those of $\mathbf{y}$ at user marked coordinates. In the case of pairwise derivative filters $\{\mathbf{F}_\gamma\} = \{[1, -1], [1, -1]^T\}$, this constrains the layer derivatives to be close to those of the input mixture at locations marked by the user. We remark that the likelihood of Equation (5.11) is riddled with local optima, and the problem remains massively inverse even under user input. Bayes' rule gives rise to $p(\mathbf{x}_1, \mathbf{x}_2 | \mathbf{y}, S_1, S_2) \propto p(\mathbf{y} | \mathbf{x}_1, \mathbf{x}_2, S_1, S_2) p(\mathbf{x}_1) p(\mathbf{x}_2)$. Besides introducing a mean field-based technique, the novelty of our formulation lies in the likelihood of Equation (5.11), which *allows for slack in the summation through an additional noise layer*. In contrast, previous work (Levin and Weiss, 2007) enforces exact summation to the input $\mathbf{y} = \mathbf{x}_1 + \mathbf{x}_2$ by setting $\mathbf{x}_2 = \mathbf{y} - \mathbf{x}_1$. This pushes existing image noise into the recovered layers. Our approach avoids this by separating an additional noise layer. To obtain an MMSE estimate, we infer the hidden layers by mean field. Figure 5.4 depicts a result: Hidden image layers are extracted from a user marked mixture corrupted by Gaussian noise. The input image has challenging details and texture exacerbated by the image noise. Nonetheless, our method separates

meaningful layers, which gives another example of the efficacy of mean field for generative image models.

5.5 CONCLUSION

This chapter studied mean field for continuous high-order MRFs, particularly those formulated with Gaussian mixture potentials. We not only addressed an increasing interest in Bayesian estimation beyond the 0/1-loss, but also computational limitations of existing, largely discrete inference algorithms. In particular, we showed fully factorized mean field updates to scale quadratically in the clique size of the prior. A broad application study in denoising, deblurring, noise estimation, and layer separation indicated that mean field exhibits a favorable combination of efficiency, versatility, and performance.

6

LOCALIZED IMAGE BLUR REMOVAL THROUGH NON-PARAMETRIC KERNEL ESTIMATION

CONTENTS

We address the problem of estimating and removing localized image blur, as it for example arises from moving objects in a scene, or when the depth of field is insufficient to sharply render all objects of interest. Unlike the case of camera shake, such blur changes abruptly at the object boundaries. To cope with this, we propose an automated sharp image recovery method that simultaneously determines blurred regions and estimates their responsible blur kernels. To address a wide range of different scenarios, our model is not restricted to a discrete set of candidate blurs, but allows for arbitrary, *non-parametric* blur kernels. Moreover, our approach does not require specialized hardware, an alpha matte, or user annotation of the blurred region. Unlike previous methods, we show that localized blur estimation can be accomplished by incorporating a pixel-wise latent variable to indicate the active blur kernel. Furthermore, we generalize the marginal likelihood technique of blind deblurring to the case of localized blur. Specifically, we integrate out the latent image derivatives to permit marginal density estimates of both blur kernels and their regions of influence. We obtain sharp images in applications to both object motion blur and defocus blur removal. Quantitative results on two novel datasets as well as qualitative results comparing to a range of specialized methods demonstrate the versatility and effectiveness of our non-parametric approach.

(a) (b)

Figure 6.1: Examples of motion blur unaligned with the image axes. (a) Fast moving vehicle at an angle (Raskar et al., 2006). (b) Rotational motion (Levin et al., 2008b).

6.1 INTRODUCTION

In many realistic conditions, images are degraded by localized image blur. For example, if limited illumination necessitates a slow camera shutter, rapidly moving (foreground) objects frequently cause motion blur. Another example is defocus blur, when image regions of interest have been rendered out of focus. In digital photography, this is often undesirable. Removing the object blur, while preserving sharp image regions, is thus an important application. The challenge in both cases is that neither the blur nor the extent of the affected region is known. While it may be possible to address these problems with user input, specialized hardware, or multiple exposures, our focus lies on automatic solutions that operate on a single image. In this chapter, we develop an integrated approach for blind removal of spatially-varying blur, which is able to determine the extent of blurred regions and simultaneously estimate the non-parametric blur kernel causing the loss of image details.

Most blind deblurring techniques focus on removing spatially uniform blur, (*e.g.*, Xu and Jia (2010)), or smoothly varying blur, such as from camera rotation (Whyte et al., 2010). These approaches cannot handle abruptly varying object blur; strong image artifacts arise when applying them nevertheless. To remove spatially localized blur, the blurry pixels must first be identified as such, while any sharp region should remain intact. This is a massively inverse problem, since unknown blurs, the locations where they apply, and the latent image all have to be estimated. Existing approaches to localized blur use a variety of constraints to regularize the problem, for example by relying on an alpha matte (Jia, 2007; Dai and Wu, 2009), thus on user interaction, or by employing modified hardware (Martinello and Favaro, 2011). In contrast, we develop an automatic approach that requires only a single image as input. Other existing methods for recovering localized blur provide additional constraints by choosing a likely blur from a

predefined candidate set. For example, in the case of motion blur, these are often box filters with known orientation (Levin, 2007; Chakrabarti et al., 2010; Couzinié-Devy et al., 2013), typically horizontal and vertical. However, image blur is not always perfectly aligned with the image axes (see Figure 6.1); restricting the kernel to axis-aligned motion can even cause strong deblurring artifacts (Figure 6.6(c)).

To address this limitation, we propose a novel approach for estimating and subsequently removing *localized*, *non-parametric* blurs of any type. In particular, we develop a probabilistic blur model, whereby each pixel is augmented with a latent variable that indicates which blur kernel is active at that site. To regularize the problem, we use a spatial prior on these latent indicators modeling the coherence of realistic objects and background areas. Moreover, we robustly infer the indicator variable configuration and blur kernels using the well-proven variational Bayes framework, by integrating over the latent image derivatives in a novel, generalized version of the marginalized maximum a-posteriori (MAP) approach (Levin et al., 2011). On the one hand, this allows to identify the different image areas that are affected by each blur (see Figure 6.8), and on the other, to estimate arbitrary, non-parametric blur kernels (Figure 6.3(b), 6.5(b)). We evaluate our joint estimation approach both quantitatively and qualitatively. First, we analyze the basic properties of our method, showing that it performs almost as well as if the ground-truth blur locations were given, despite inferring many more unknowns. Second, we show that our approach can cope with realistic cases of object blur, in which the image blur is not perfectly horizontal or vertical, but may slightly deviate from the axes (as it occurs, *e.g.*, in Figure 6.1). This is unlike most previous work that assumes perfectly axis-aligned blur to keep the hypothesis space manageable. Further, qualitative results and comparisons to other methods for object motion deblurring and defocus blur removal illustrate the versatility of our approach. For example, we show that our method can simultaneously cope with motion and defocus blur when these occur in a single image.

6.2 RELATED WORK

There is a large body of literature on blind deblurring—the problem of recovering a sharp image from a blurry input without knowledge of the blur kernel causing the image degradation (see Kundur and Hatzinakos (1996a); Wang and Tao (2014) for overviews of standard techniques). Variational Bayesian inference was recognized early on as an effective algorithm to cope with the ill-posedness inherent to blind deblurring (Miskin and MacKay, 2000; Fergus et al., 2006). Other work followed suit (Whyte et al., 2010; Levin et al., 2011; Wipf and Zhang, 2013). We adopt this algorithm here as well, as it is theoretically well-founded and performs well in practice (Levin et al.,

2011; Wipf and Zhang, 2013). However, our work extends and improves on previous variational deblurring algorithms by estimating not only blur kernels, but also a map of pixels each kernel acts upon. The majority of deblurring approaches work under the assumption of a uniformly blurred image (*e.g.*, Cho and Lee (2009); Levin et al. (2011)). In a strict sense, however, the assumption of spatially uniform blur is almost always violated in practice. For example, in the case of camera shake, a rotational motion component can cause the blur kernel to vary smoothly over the image, which requires a more accurate model of the image formation process (Whyte et al., 2010). In contrast to camera shake, the primary focus of this chapter lies on blur arising from independent object motion in a static scene (Figure 6.7), or from defocused objects at a certain depth in the scene (Figure 6.5). This type of image blur exhibits abrupt, rather than smooth, spatial changes within the image, and we refer to it as *localized blur*.

To mitigate the ill-posedness of localized blur analysis, more than one input image can be used (Bar et al., 2007). Alternatively, hardware approaches include a fluttered shutter (Raskar et al., 2006), aperture patterns (Levin et al., 2007; Martinello and Favaro, 2011), or camera motion during exposure (Levin et al., 2008b). We here focus on the purely *image-based* setting with only a single input image, as it is characteristic of the majority of usage scenarios. This challenging inverse problem has been regularized through user assistance (Jia, 2007; Dai and Wu, 2009). Here, the user marks the blurred object by brush strokes, so that the corresponding alpha matte can be extracted and used for further processing. In contrast, we propose to identify the blurred pixels without user supervision in this chapter. Our automatic algorithm is based only on the raw pixel information of a single, standard camera image. Methods of this type (Levin, 2007; Chakrabarti et al., 2010; Couzinié-Devy et al., 2013; Kim et al., 2013) often limit the space of blur kernels, or they do not treat the case of defocus blur. In this chapter, we put forward a novel, Bayesian model of localized image blur, which incorporates a latent variable to switch pixel-wise between different blur kernels. We demonstrate that our model obtains high quality results, yet is flexible enough to, for example, permit removing both defocus and motion blur even in the case when these occur simultaneously in a single image.

6.3 LOCALIZED BLUR MODEL

In contrast to the spatially uniform blur case, we here allow for several blur kernels $\mathbf{k} = \{\mathbf{k}_i\}$ to act upon disjoint regions of the image. To express this formally, we augment each blurry pixel y_n with a latent indicator variable $\mathbf{h}_n$. Each $\mathbf{h}_n \in \{0,1\}^M$ is a binary, unit-sum vector indicating the blur kernel that is active at the n-th pixel. Denoting

the set of latent variables for all pixels as $\mathbf{h} = \{\mathbf{h}_n\}$, we express the likelihood of an image $\mathbf{y}$ under spatially-varying blur as

$$p(\mathbf{y}|\mathbf{x}, \mathbf{h}, \mathbf{k}) = \prod_n \prod_i \mathcal{N}\left(y_n|\mathbf{k}_i \otimes \mathbf{x}_n, \sigma^2\right)^{h_{ni}}, \tag{6.1}$$

where $\mathbf{x}_n$ denotes the n-th clique of the sharp image that, under convolution, gives rise to a single blurry pixel y_n. Here, σ^2 denotes the variance of additive Gaussian noise used to model the fluctuations in the imaging process. In practice, we estimate the blurs and indicator variables in the gradient domain, such that the likelihood is

$$p(\mathbf{y}|\partial_x, \mathbf{h}, \mathbf{k}) = \prod_{n,i} \left[\prod_j \mathcal{N}\left((\partial_j y)_n|\mathbf{k}_i \otimes (\partial_j \mathbf{x})_n, \sigma^2\right) \right]^{h_{ni}}. \tag{6.2}$$

While our localized blur model can, in principle, be used with any number of blurs, we restrict ourselves to the case of $M = 2$ kernels here. In most experiments, we fix one kernel as the δ (identity) filter to explain the sharp regions. This is accurate enough for many images, which often contain only a single, blurred region of interest (ROI). We used this approach for all experiments except Figure 6.9.

Estimating spatially-varying blur is a massively inverse problem and thus requires incorporating appropriate prior knowledge. As is usual in the blind deblurring literature, we first rely on an image prior on the latent image $\mathbf{x}$. We specifically use a gradient prior, modeled as a Gaussian scale mixture (GSM), *i.e.*, a weighted sum of zero-mean normal distributions $\overline{p}(u) = \sum_l \pi_l \mathcal{N}(u|0, \sigma_l^2)$, where the positive weights π_l sum to unity. We capture the derivative statistics of natural images by fitting a GSM to a characteristic, heavy-tailed histogram of derivative responses arising of 200 images extracted from the BSDS500 dataset. The resulting gradient prior is

$$p(\nabla \mathbf{x}) = p(\partial_1 \mathbf{x}) \cdot p(\partial_2 \mathbf{x}) = \prod_j \prod_n \overline{p}\left((\partial_j x)_n\right). \tag{6.3}$$

Note that the GSM parameters (π_l, σ_l) are spatially invariant across the image. Such an image prior alone is not sufficient for reliably recovering the blur and estimating where in the image it occurs. To address that we observe that blur degradation tends to occur in connected regions (object motion or defocus blur). We encode this prior knowledge on the indicator variables $\mathbf{h}$ with a pairwise Markov random field (MRF), specifically using a Potts prior

$$p(\mathbf{h}) \propto \prod_{(l,m)\in\mathcal{N}} \exp\left(-\lambda \cdot [\mathbf{h}_l \neq \mathbf{h}_m]\right), \tag{6.4}$$

where $\mathcal{N}$ denotes a set of pairwise, neighboring cliques (in the experiments, 8-neighborhood), λ is a regularization weight, and $[\cdot]$ denotes the Iverson bracket.

To estimate the blur kernels and their spatial extent, we adopt a Bayesian approach and formulate the posterior over the unknowns as

$$p(\nabla \mathbf{x}, \mathbf{h}, \mathbf{k}|\mathbf{y}) \propto p(\mathbf{y}|\nabla \mathbf{x}, \mathbf{h}, \mathbf{k}) p(\nabla \mathbf{x}) p(\mathbf{h}). \tag{6.5}$$

This is a significantly more challenging inverse problem than uniform blind deblurring, which only consists of estimating a single blur kernel to explain the input image. In contrast, we here need to not only infer blur kernels, but also a plausible configuration of the latent indicator variables. Once blur kernels and indicator variables are determined, we recover the actual intensities of the desired sharp image in a non-blind deblurring step as detailed below.

6.4 INFERENCE

Building upon the robust and well-proven marginalized MAP framework to deblurring (Levin et al., 2011), our goal is to infer the blur kernels $\mathbf{k}$ and latent variables $\mathbf{h}$ by maximizing the densities

$$p(\mathbf{k}|\mathbf{y}) = \int p(\nabla \mathbf{x}, \mathbf{h}, \mathbf{k}|\mathbf{y}) \, d\nabla \mathbf{x} \, d\mathbf{h}, \quad \text{and} \tag{6.6}$$

$$p(\mathbf{h}|\mathbf{y}) = \int p(\nabla \mathbf{x}, \mathbf{h}, \mathbf{k}|\mathbf{y}) \, d\nabla \mathbf{x} \, d\mathbf{k}. \tag{6.7}$$

Since exact inference is intractable, we use variational Bayesian (mean field) approximate inference (Minka, 2005). To fulfill the necessary exponential family constraint, the GSMs $\overline{p}$ of the gradient prior must therefore be augmented with latent mixture coefficients. For this we introduce a binary-valued, unit-sum vector $\mathbf{v}$ such that $\overline{p}(\mathbf{v}) = \prod_l \pi_l^{v_l}$, and formulate

$$\overline{p}(u|\mathbf{v}) = \prod_l \mathcal{N}(u|0, \sigma_l^2)^{v_l}. \tag{6.8}$$

The length of $\mathbf{v}$ equals the number of GSM components (we used 13). Performing this common expansion (*e.g.*, Levin et al. (2011)) for each pixel n of each derivative j yields vectors $\mathbf{v}_{nj}$, which we summarize as $\mathbf{t} = \{\mathbf{v}_{nj}\}$. This augmentation preserves the original gradient prior, *i.e.* $\sum_{\mathbf{t}} p(\partial_x, \mathbf{t}) = p(\partial_x)$, but allows to conveniently approximate the expanded posterior

$$p(\nabla \mathbf{x}, \mathbf{h}, \mathbf{k}, \mathbf{t}|\mathbf{y}) = p(\mathbf{y}|\nabla \mathbf{x}, \mathbf{h}, \mathbf{k}) p(\nabla \mathbf{x}, \mathbf{t}) p(\mathbf{h}) \tag{6.9}$$

by a tractable, fully-factorized density

$$q(\nabla \mathbf{x}, \mathbf{h}, \mathbf{k}, \mathbf{t}) = q(\nabla \mathbf{x}) q(\mathbf{h}) q(\mathbf{k}) q(\mathbf{t}) \tag{6.10}$$

using variational Bayesian inference. The marginals $q(\mathbf{h})$ and $q(\mathbf{k})$ of the approximate density then serve as surrogates for the actual

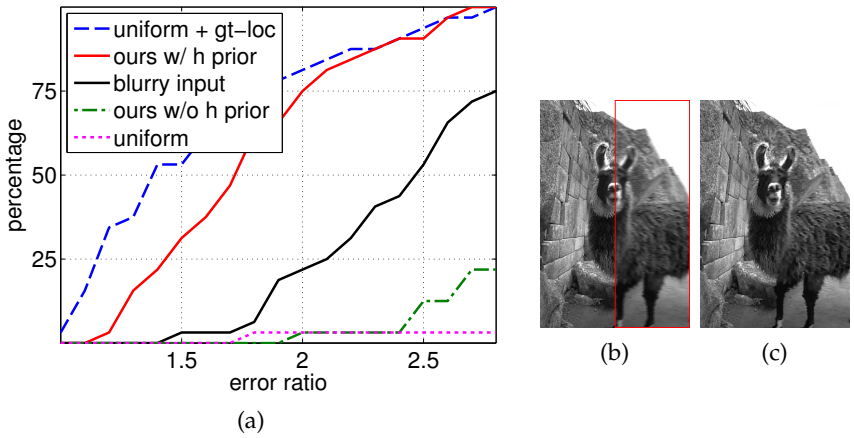


Figure 6.2: Quantitative evaluation on BSDS images. (a) Cumulative histogram of error ratio reporting the percentage of test instances with an error ratio below a certain value. Despite estimating the blur location, our method (*ours w/ h prior*) performs close to blind deblurring with known ground-truth blur location (*uniform + gt-loc*). Standard uniform deblurring or not using the indicator prior (*ours w/o h prior*) performs much worse. (b) Example input image with blurred sub-region marked by a red rectangle. (c) Our deblurring result. The images are best viewed by zooming in using a computer display.

marginals $p(\mathbf{h}|\mathbf{y})$ and $p(\mathbf{k}|\mathbf{y})$ of the posterior. As the update steps of variational Bayesian inference are somewhat involved, we include them in Appendix B.

We use coarse-to-fine estimation to aid and accelerate convergence, which is a standard approach to overcome ill-posedness in blind deblurring. Note that in our case the problem is even more difficult due to the need to distinguish blurred from sharp pixels. At each scale s, we run variational inference to fit an approximate density $q(\partial_x^s, \mathbf{h}^s, \mathbf{k}^s, \mathbf{t}^s)$ to the posterior $p(\partial_x^s, \mathbf{h}^s, \mathbf{k}^s, \mathbf{t}^s|\mathbf{y}^s)$; for ease of notation, we omit the scale index for probability densities, *i.e.* $p \equiv p^s$, $q \equiv q^s$. When moving to the next finer level $s-1$, we initialize the new indicator and kernel distributions $q(\mathbf{h}^{s-1})$ and $q(\mathbf{k}^{s-1})$ by resizing and interpolating the parameters of $q(\mathbf{h}^s)$ and $q(\mathbf{k}^s)$. The multiscale variational inference procedure yields indicator and kernel densities $q(\mathbf{h}^0)$ and $q(\mathbf{k}^0)$ at the finest level, from which we obtain the final estimates

$$\mathbf{h} = \operatorname{argmax} q(\mathbf{h}^0), \quad \mathbf{k} = \operatorname{argmax} q(\mathbf{k}^0), \tag{6.11}$$

which generalizes marginalized MAP deblurring (Levin et al., 2011).

Let us briefly consider the intricacies of double blur initialization at the highest level of the pyramid, where the kernels are of small size (*e.g.*, 7×7). In case one blur is held fixed as the identity kernel throughout the procedure, the other may also be initialized as identity. This is done for all experiments except Figure 6.9, where two

(a) (b)

	PSNR	SSIM	MAE
Blurry input	25.59	0.816	6.13
Our algorithm	26.37	0.835	6.04
Uniform (Levin et al., 2011)	18.73	0.562	20.80
Chakrabarti et al. (2010) & non-blind	26.21	0.818	6.24

(c)

Figure 6.3: Synthetic motion deblurring of VOC objects. (a) Example motion blurred image with ground truth blur in top right corner. [While the foreground object is from the VOC database, the background is a resized version of Wikimedia Commons file "Kooperative kleinräumige Nischenvielfalt.jpg" by EwigLernender, licensed under CC BY-SA 3.0.] (b) Our deblurring result with estimated kernel in top right corner. (c) Average values over 10 motion blurred images. The first row specifies the quality of the blurry input images. Values printed in red are worse than those of the input images.

varying kernels are estimated. These must be initialized differently, as the variational update pattern (see Appendix B) would otherwise result in identical blur estimates. For the image in Figure 6.9, we have tested that the procedure succeeds under initialization with (1) δ and Gaussian blur, and (2) horizontal box filter and Gaussian blur, with the latter producing the better, depicted result. Triple blur estimation with horizontal and vertical box filters, as well as Gaussian blur as initial values also succeeds. Future work must determine which starting values are optimal over a range of images. Intuition may be gained from the initialization of EM clustering (Gupta and Chen, 2011).

From the inferred indicator variables, we can look at a particular slice $\mathbf{l}_i = \{h_{ni}\}$, which is a binary image labeling where a value of 1 indicates that the i-th blur kernel is active at a pixel. Figure 6.8 shows an example of a pixel labeling inferred by our approach. Note that the labelings allow to cope with partial occlusions of blurred image regions (Figures 6.5, 6.9). After blur kernels and blur maps have been determined, they can be used to remove the blur in a non-blind deblurring step. In the common case of a single blur restricted

(a) (b) (c)

Figure 6.4: Comparison to user assisted removal of spatially-varying blur. (a) Input (Dai and Wu, 2009). (b) Our result. (c) Result of Dai and Wu (2009). Note that the latter method requires the user to mark the blurred object, while our algorithm is automatic. The images are best viewed by zooming in using a computer display.

to a partial region of an otherwise sharp image, we formulate the data term $E_{data} \equiv E_{data}(\mathbf{x}, \mathbf{h}, \mathbf{k}, \mathbf{y})$ for non-blind deblurring as

$$E_{data} = \|\mathbf{y} - \mathbf{k}_1 \otimes (\mathbf{l}_1 \odot \mathbf{x}) - (1 - \mathbf{k}_1 \otimes \mathbf{l}_1) \odot \mathbf{x}\|^2, \tag{6.12}$$

which for both motion and defocus blur accurately models the transparency at the object boundary (Köhler et al., 2013). In the general case of more than one non-trivial blur (different from the delta kernel), we use $E_{data} = \|\mathbf{y} - \sum_i \mathbf{k}_i \otimes (\mathbf{l}_i \odot \mathbf{x})\|^2$. We then recover the sharp image by minimizing

$$E_{data}(\mathbf{x}, \mathbf{h}, \mathbf{k}, \mathbf{y}) + \gamma \sum_{j,n} |(\partial_j x)_n|^{0.8} \tag{6.13}$$

w.r.t. $\mathbf{x}$ using iteratively re-weighted least-squares. The objective function in Equation (6.13) combines the data term with a sparsity prior on the image derivatives weighted by $\gamma > 0$. The term $|\cdot|^{0.8}$ is chosen as a robust penalty function (Levin et al., 2007), while in practice, the weight γ should be adjusted to the magnitude of image noise. Note that in regions unaffected by any blur, minimizing the energy (6.13) simply corresponds to denoising.

6.5 EXPERIMENTAL EVALUATION

6.5.1 *Quantitative Experiments*

CUMULATIVE HISTOGRAM OF ERROR RATIO. To quantitatively measure our model's capacity to identify image blur, we use a data set consisting of 32 BSDS500 images in which a sub-region has been synthetically blurred by one of 8 different box filters; the remainder of the test image was left sharp. We rely on the SSD error ratio SSD_{est}/SSD_{gt} (Levin et al., 2011) as performance metric, adapted to our context as the ratio between SSD error after deblurring with the estimated values for kernel and pixel labeling (SSD_{est}), and SSD error

(a) (b)

Figure 6.5: Defocus blur removal. (a) Input image with unfocused background occluded by an in-focus foreground object (Zhang and Cham, 2009). (b) All-focus result: Our algorithm automatically sharpens the blurred background while preserving foreground pixels. The estimated defocus blur is shown in the top left corner.

after deblurring with the ground-truth values for kernel and labeling (SSD_{gt}). Figure 6.2 displays a cumulative histogram of error ratio on the dataset. We make several observations: (1) We compare to a state-of-the-art uniform deblurring method (Levin et al., 2011) that has been applied only to the blurry region, thus assumes knowing the extent of the blurry region in advance. Despite solving a much harder problem, our approach performs quite close to this impractical upper bound. (2) If we apply the same uniform method to the entire image instead of limiting it to the extent of the blurred region, the image quality sharply deteriorates, even below the level of the input image. A correct labeling of the regions affected by blur is thus crucial. (3) Comparing to a variant of our technique that does not rely on the spatial Potts prior on the indicator variables shows that this prior knowledge is a key factor in our algorithm's success.

REALISTIC MOTION BLUR. To understand the benefits of our nonparametric approach to localized blur, we first study the orientations of realistic motion blur. We cropped motion blurred patches from 94 images of a real motion blur dataset (Shi et al., 2014), and estimated a 51×51 blur kernel on each patch using a uniform blind deblurring method (Xu and Jia, 2010). We then matched every estimated blur to one of 180 candidate orientations by computing the correlation values (up to shifts) of angled box filters with the estimated blur and then choosing the orientation with the highest score. Of the measured orientations, 61% are tilted away from horizontal or vertical, while 86% lie in a range of $\pm 20°$ around the axes. Based on these observations, we designed a test data set with ground truth by simulating images affected by object motion blur. In particular, we extracted foreground objects from images of the VOC data set using

(a) (b) (c)

Figure 6.6: Motion deblurring. (a) Motion blurred input image (frame from the *face* sequence of Portz et al., 2012). (b) Our result. (c) Result of Chakrabarti et al. (2010) + non-blind. Our algorithm removes the foreground motion blur without harming the integrity of the background.

the given ground-truth object segmentation. The motion blurred object is then inserted over a realistic, static background image. This is done by warping the object along a linear blur trajectory and alpha matting it onto the background at single pixel intervals. The orientation of the blur trajectories is sampled uniformly from an interval of $\pm 20°$ around the horizontal and vertical axes, with the sample being accepted if physically plausible. The final simulated image is then obtained by averaging these frames. We created 10 such images as test data. Figure 6.3 shows an instance of the dataset together with deblurring results on ten images as measured by average peak signal-to-noise ratio (PSNR), structural similarity index (SSIM), and mean absolute error (MAE). The table shows that our approach clearly outperforms a recent high-grade motion blur segmentation and kernel estimation method (Chakrabarti et al., 2010) when combined with the same non-blind deblurring algorithm as for our model. As can be expected from the previous results, uniform deblurring (Levin et al., 2011) also fails on this data set. These results thus demonstrate that (1) to cope with realistic blur scenarios, commonly used restrictions to the space of possible blur kernels (*e.g.*, Chakrabarti et al. (2010)) should be avoided, and (2) that identifying the extent of the blur is very important for high image quality.

6.5.2 *Qualitative Results*

Figures 6.4–6.9 show results on several instances of real, blurred images from other publications. Note that for these qualitative results, the free parameters (such as the kernel size) were calibrated per image to showcase optimal performance. This is common practice in blind deblurring (Köhler et al., 2012).

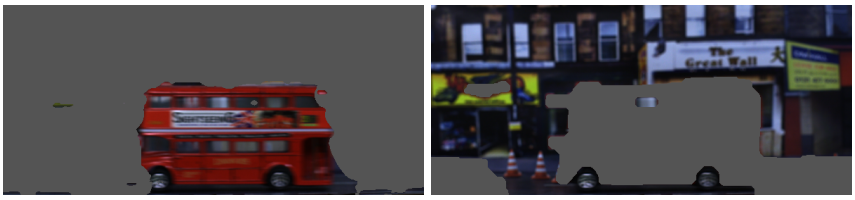
In Figure 6.4, we compare our automatic algorithm to user guided defocus removal (Dai and Wu, 2009). Here, the user manually marks the blur degraded ROI, which permits to extract a complete alpha matte of the blurred object. This naturally facilitates kernel estima-

(a) (b) (c) (d)

Figure 6.7: Motion deblurring. (a) Input image *jogger* from the database of Chakrabarti et al. (2010). (b) Our deblurring result. (c), (d) Image details before and after removal of motion blur.

tion and boundary handling. Nevertheless, we observe that our *automatic* procedure yields a visually pleasing result of similar quality. Figure 6.5 depicts a further instance of real defocus blur removal, where the blurred background is partially occluded by a foreground object. Our method successfully restores detail to the background while leaving the foreground object sharp, all without any user assistance. Figure 6.6 displays an example of motion blur as it may occur with a low-grade webcam. We here observe that our approach is able to counteract the motion blur while leaving the sharp image region untouched. In comparison, the localized blur estimation method of Chakrabarti et al. (2010) does not cope with the motion blur nearly as well as our approach. Figure 6.7 shows another instance of successful motion blur removal.

Figure 6.9 contains a particularly challenging instance of spatially-varying blur: The foreground object is blurred by motion, while the background is out of focus. Figure 6.8 shows that our algorithm correctly identifies the motion blurred pixels, while on the other hand, a recent motion blur segmentation method (Chakrabarti et al., 2010) erroneously labels the unfocused image background as motion blurred. In Figure 6.9, we can moreover observe that our algorithm succeeds in automatically sharpening both motion and defocus blurred regions using distinct blur kernel estimates for each automatically identified region. For comparison, we include a deblurring result obtained using a camera aperture designed specifically for the purpose of removing motion and defocus blur (Martinello and Favaro, 2011). Despite being independent of any dedicated hardware, our approach achieves at least competitive results, since the defocused background does not



(a) (b)

Figure 6.8: Motion blur detection. (Figure 6.9 shows the blurry input image). (a) Pixels labeled as motion blurred by our algorithm. (b) Motion blur segmentation of Chakrabarti et al. (2010). Note that in the latter instance, the defocused region is wrongly labeled as motion blurred. See Figure 6.9 for our deblurring result.

suffer from over-sharpening artifacts and the result from the motion blurred object is sharper and contains fewer artifacts.

6.5.3 *Runtime*

With regard to computational effort, we measured runtimes on the VOC data set (Figure 6.3) consisting of 10 color images with an example size of 500×375. On average, a MATLAB implementation of our algorithm took 4.7 minutes, comprising 3.5 minutes for estimating the localized blur, and 1.2 minutes for the final non-blind deblurring step. Note that our framework is very general, since kernels are estimated from a real-valued, non-parametric space, allowing for many different cases including object motion, camera shake, and defocus blur. To further put this into context, we measured 2 minutes on average for the uniform baseline (Levin et al., 2011), which also uses variational inference. A specialized algorithm using a candidate set of just 24 axis-aligned motion blurs (Chakrabarti et al., 2010) required 1.8 minutes on average (1 minute for localized blur estimation, and 50 seconds for non-blind deblurring). However, the restoration performance is significantly worse than our method by 0.16/0.017/0.20 in PSNR/SSIM/MAE. Measurements were made on a machine with a 3.20GHz Core i7 3930K processor.

6.6 CONCLUSION

We considered the problem of estimating and removing localized object blur, which exhibits sudden changes across the image plane. To address this, we used a new Bayesian formulation that incorporates pixel-wise latent variables indicating which blur kernel is active. Our approach generalizes marginalized MAP and allows estimating non-parametric blurs, instead of limiting the kernels to a discrete candidate set. Quantitative experiments showed that our approach allows to better cope with motion blurs tilted around the image axes, as we

(a)

(b)

(c)

Figure 6.9: Simultaneous removal of motion and defocus blur. (a) Top: A complex scene having both motion and defocus blur (Martinello and Favaro, 2011). (b) Our deblurring result. (c) The result of Martinello and Favaro (2011). Note that the latter method relies on a customized camera aperture, while our algorithm is applicable to off-the-shelf camera images. Our approach better recovers the motion blurred bus.

found to occur frequently in practice. High-quality instances of real motion and defocus blur removal demonstrate the effectiveness of our technique. Our non-parametric framework can successfully handle blurs of very different types, and we demonstrated results with performance competitive to user- or hardware-assisted techniques, despite our method being automatic.

7

INTERLEAVED REGRESSION TREE FIELD CASCADES FOR BLIND IMAGE DECONVOLUTION

CONTENTS

Image blur from camera shake is a common cause for poor image quality in digital photography, prompting a significant recent interest in image deblurring. The vast majority of work on blind deblurring splits the problem into two subsequent steps: First, the blur process (*i.e.*, blur kernel) is estimated; then the image is restored given the estimated kernel using a non-blind deblurring algorithm. Recent work in non-blind deblurring has shown that discriminative approaches can have clear image quality and run-time benefits over typical generative formulations. In this chapter, we propose a cascade for blind deblurring that alternates between kernel estimation and discriminative deblurring using Regression Tree Fields (RTFs). We further contribute a new dataset of realistic image blur kernels from human camera shake, which we use to train the discriminative component. Extensive qualitative and quantitative experiments show a clear gain in image quality by interleaving kernel estimation and discriminative deblurring in an iterative cascade.

7.1 INTRODUCTION

Camera shake causes light quantities of several, spatially distinct locations of the scene to coincide at a single coordinate of the image plane during exposure. Modern cameras stabilize the lens or the sensor, but this can only counteract relatively small camera motion. Besides limiting the user experience in consumer digital photography, image blur from camera shake is also encountered in scientific and

(a) Blurry (b) Deblurred

Figure 7.1: Deconvolution with interleaved RTF cascade. Blurred input from Jia (2013).

industrial applications, causing wide interest in removing the effects of such blur (Wang and Tao, 2014).

The most widely adopted restoration approach is to first estimate the blur kernel (Fergus et al., 2006; Joshi et al., 2010; Xu and Jia, 2010; Whyte et al., 2010; Levin et al., 2011; Krishnan et al., 2011; Hirsch et al., 2011; Xu et al., 2013), often by making some statistical assumptions on the unknown sharp image. In a separate, *non-blind* step the sharp image is then restored given the kernel estimate, which is *held fixed during the procedure*. Many modern non-blind deblurring algorithms adopt a *generative* approach and impose prior knowledge on the image (Krishnan and Fergus, 2009; Schmidt et al., 2011; Zoran and Weiss, 2011). While accurate generative models exist, *e.g.*, high-order Markov random fields (MRFs) (Schmidt et al., 2011), their extensive computational demands prohibit the use as part of the kernel estimation phase. The origins of our work lie in recent *discriminative* approaches to non-blind deblurring, which use a neural network (Schuler et al., 2013), or prediction cascades of RTFs (Schmidt et al., 2013) or shrinkage fields (Schmidt and Roth, 2014). Their benefit is that they deliver high-quality image estimates, which outperform most generative approaches, at a fraction of the computational cost. However, their use in kernel estimation or blind deblurring has not been considered so far.

In this chapter, we propose an *RTF cascade for blind deblurring*, which alternates between discriminative deblurring and re-estimating the blur kernel using the refined image prediction. This generalizes previous work on RTF cascades for non-blind deblurring (Schmidt et al., 2013) to the blind deblurring task. One feature of discriminative deblurring approaches is that typical errors made by the kernel estimation procedure are learned during training, so they can be compensated for in the image recovery procedure. Kernel estimation and non-blind deblurring are therefore trained as inter-related components. To the best of our knowledge, we are the first to use discriminative image

prediction custom trained to the blur updates of a blind deconvolution procedure.

To train a powerful discriminative model for image restoration, the training data should cover a rich variety of camera motions to avoid overfitting. We address this by further contributing a novel dataset of blur kernels obtained by photographing an isolated point light source under human camera shake, and use this data to train our model. We evaluate our approach extensively, and find that it clearly outperforms other recent methods from the literature.

7.2 RELATED WORK

An early approach for camera shake removal was proposed by Fergus et al. (2006), using a variational Bayesian approach. Later research (Levin et al., 2011; Wipf and Zhang, 2013) showed that estimating the blur kernel by (approximately) marginalizing over the latent sharp image allows to cope with the ill-posed nature of the problem and also yields high-quality results in practice. Maximum a-posteriori (MAP) approaches to kernel estimation typically excel in terms of fast running time (Cho and Lee, 2009). However, a naïve implementation is likely to favor the trivial no-blur solution (Levin et al., 2009). Fortunately, this can be circumvented by intermediate shock or bilateral filtering of the latent image (Xu and Jia, 2010; Cho and Lee, 2009), alternating minimization schemes (Perrone and Favaro, 2014), or clever design of the image prior (Krishnan et al., 2011; Xu et al., 2013; Shan et al., 2008; Michaeli and Irani, 2014).

Blind deblurring algorithms often rely on first detecting a set of useful image edges, from which the blur kernel can be estimated robustly (Joshi et al., 2008; Sun et al., 2013). In particular, Xu and Jia (2010) estimate the blur on salient edge locations and enforce the expensive kernel sparsity constraint only once, at the end of the multi-scale blur estimation procedure. Since this method yields a favorable combination of efficiency and performance (see, *e.g.*, Köhler et al. (2012)), we use it to bootstrap our approach. However, note that our framework can also operate over multiple image scales, taking a delta kernel as initial input. An alternative to image-based blur estimation is to use motion sensor data recorded during exposure to reconstruct the kernel (Joshi et al., 2010). Another technique to boost restoration performance is to use context-specific sharp image examples (Sun et al., 2014). We focus here on the more common post-capture scenario, where only the blurry image is given.

Discriminative approaches to image restoration often take the form of conditional random fields (CRFs). Due to their computational advantages, Gaussian CRFs have attracted particular attention. Tappen et al. (2007) were among the first to propose discriminatively trained Gaussian CRFs. A more recent variant is termed Regression Tree

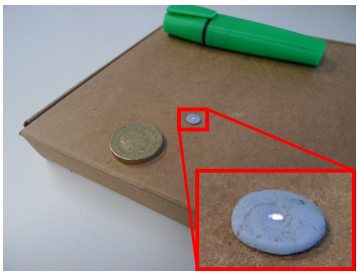


Figure 7.2: White LED point light source setup.

Fields (RTFs) (Jancsary et al., 2012a). The parameters of these Gaussian CRFs are determined by non-parametric regression trees; we provide more technical details on their application to image deblurring in Section 7.4. RTFs have proven effective in a variety of restoration tasks, including denoising, inpainting, and colorization (Jancsary et al., 2012a,b). Recently, two different kinds of discriminative non-blind deblurring approaches have been proposed: (1) using a neural network (Schuler et al., 2013), and (2) based on stacked CRF *cascades* (Schmidt et al., 2013; Schmidt and Roth, 2014). We here choose to rely on RTF cascades (Schmidt et al., 2013), since they do not require the test-time blur kernel to be known at training time, which is a prerequisite for using them as a component in a blind deblurring approach. In our work, we further explore the capacity of discriminative cascades by generalizing them to blind image deblurring through interleaving the discriminative prediction stages with blur kernel estimation.

To generalize well, a discriminative model must be exposed to a sufficiently large variety of training data. However, publicly available blurs resulting from real camera shake are limited to 8 instances in the dataset of Levin et al. (2009), and 12 instances from the dataset of Köhler et al. (2012). In our work, we capture realistic blurs by recording human camera shakes, and we validate the novel data by using it to train a state-of-the-art discriminative deblurring model. Note here that the recorded data is publicly available[1], and may benefit other research too, *e.g.*, generative blur modeling.

7.3 RECORDING NATURAL CAMERA SHAKE

Training a good regressor generally requires many instances of realistic data. To generate realistic blur kernels for training our RTF cascade, we recorded trajectories of a point light source under camera shake. For this we used a white light-emitting diode (LED) (OSA Opto Light Series 400 white) as light source, placed within a cardboard box. We limited the spatial area of the light source as well as the overall amount of emitted light by placing a blue tack onto the LED, then

1 https://bitbucket.org/visinf/projects-interleaved-rtf

piercing it finely with a needle, producing a white point light source of high intensity; the cardboard box is shown in Figure 7.2.

We placed the box in an entirely dark room and recorded images from approximately 2 to 4 meters distance with a handheld Panasonic Lumix DMC-LX3 CCD camera. Capturing was done in 12 bit RAW format in manual mode at different focus depths (ISO 80, 500ms exposure, F2.0). We converted the images to raw tagged image file format (TIFF) using dcraw (v9.17, with `dcraw -T -v -4 -D`), then removed the constant black level. Because of the low ISO, almost no dark current noise remained in the digitized signal. Also, there were no saturated pixels. However, because of optical dispersion and different spectral sensitivities, the four color channels in the RAW frame (R, G1, G2, B) had different intensities and spatial blur. Thus, the RAW RGGB signal resembled a checkerboard pattern that could not be removed by applying a scalar gain factor to each channel. Because the green channels are the most sensitive, we simply used the G1 channel and discarded the other channels. We centered and normalized the blur kernel. Note that we did not observe any aliasing artifacts in the obtained blur kernels. Figure 7.3 shows examples of recorded camera shakes. Overall, we generated 192 blur kernels. Note that this data set captures the physical process and human aspects of camera shake. Future research could involve recording spatially varying blur using a grid of LED point light sources.

Figure 7.3: Instances of realistic blur kernels used for model training (Section 7.4.3). The blurs were obtained by recording the trajectory of a point light source under human camera shake.

7.4 BLIND DECONVOLUTION CASCADES

7.4.1 *Standard Non-blind RTF Cascades*

As is most common, we model the formation process of image blur as convolution under additive noise, $\mathbf{y} = \mathbf{k} \otimes \mathbf{x} + \mathbf{n}$. Thereby, $\mathbf{y}$ denotes the blurry input image, $\mathbf{k}$ the blur kernel, $\mathbf{x}$ the unknown sharp image, and $\mathbf{n}$ the additive noise. Specifically, we follow the standard assumption of normally distributed, white noise $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$. However, we could use a more realistic noise model such as Foi et al. (2008) as well. Solving for the sharp image given the blur kernel is an ill-posed, difficult problem. This is partly due to sensor noise being amplified by simply inverting the kernel. Furthermore, the inverse is not properly

defined if the blur kernel contains zero frequencies. Therefore, it is necessary to impose additional knowledge. In contrast to the many generative approaches to deconvolution, we here choose a recent, discriminative framework for image recovery, namely Regression Tree Fields (RTFs) (Jancsary et al., 2012b), to model the parameters of the posterior probability $p(\mathbf{x}|\mathbf{y},\mathbf{k})$ directly.

RTFs are Gaussian CRFs that derive their expressiveness from inferring the parameters of the local potentials using regression trees acting locally on input image features. Each tree stores at its leaves a linear term and precision matrix to define the quadratic energy contribution from the local factor variables. Regressing the potential parameters allows to overcome the apparent simplicity of Gaussian potentials, while taking full advantage of their inherent efficiency. Note that both the regression trees and the potential parameters stored at the leaves are learned in a principled, joint fashion to minimize a loss function (here, negative peak signal-to-noise ratio (PSNR) on the training data. For more details on RTFs, we refer to Jancsary et al. (2012a,b).

An RTF model for deblurring can be formulated as a Gaussian CRF of the form

$$p(\mathbf{x}|\mathbf{y},\mathbf{k}) \propto \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}(\mathbf{y},\mathbf{k}),\mathbf{C}(\mathbf{y},\mathbf{k}))\,, \tag{7.1}$$

whereby the parameters of the mean $\boldsymbol{\mu}(\mathbf{y},\mathbf{k})$ and covariance matrix $\mathbf{C}(\mathbf{y},\mathbf{k})$ are partly regressed from the input image $\mathbf{y}$ by the RTF framework, with the blur kernel $\mathbf{k}$ being held fixed as a constant. In more detail, let $\mathbf{T}_\mathbf{k}$ denote the Toeplitz matrix expressing convolution by blur kernel $\mathbf{k}$, such that the identity $\mathbf{T}_\mathbf{k}\mathbf{x} \equiv \mathbf{k} \otimes \mathbf{x}$ is fulfilled. Motivated by generative approaches to deblurring, Schmidt et al. (2013) show that the covariance and mean of the Gaussian CRF in Equation (7.1) may be chosen as

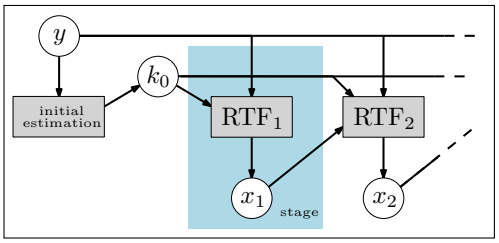
$$\mathbf{C}(\mathbf{y},\mathbf{k}) = \left(\mathbf{W}(\mathbf{y}) + \frac{1}{\sigma^2}\mathbf{T}_\mathbf{k}^\top\mathbf{T}_\mathbf{k}\right)^{-1} \tag{7.2}$$

$$\boldsymbol{\mu}(\mathbf{y},\mathbf{k}) = \mathbf{C}(\mathbf{y},\mathbf{k})\left(\mathbf{w}(\mathbf{y}) + \frac{1}{\sigma^2}\mathbf{T}_\mathbf{k}^\top\mathbf{y}\right), \tag{7.3}$$

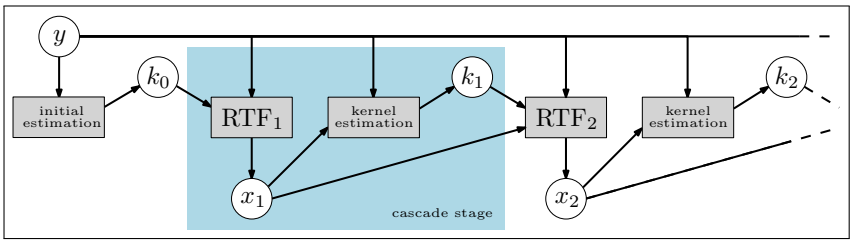
whereby the matrix $\mathbf{W}(\mathbf{y})$ and vector $\mathbf{w}(\mathbf{y})$ are regressed from the input image by the RTF framework. Overall, inference consists of regressing the CRF parameters and subsequently computing the prediction as $\operatorname{argmax}_\mathbf{x} p(\mathbf{x}|\mathbf{y},\mathbf{k}) = \boldsymbol{\mu}(\mathbf{y},\mathbf{k})$.

However, it is not easy to regress optimal potential parameters from the input image immediately, because the blur strongly obfuscates the image content, for example by creating ghosting-like overlays of edges in uniform regions. This can be overcome by stacking RTFs into a cascade (Schmidt et al., 2016), which generates a sequence of iteratively refined sharp image estimates $(\mathbf{x}_1,\ldots,\mathbf{x}_N)$. At each level of the cascade, the corresponding RTF parameters are regressed not

only from the input image, but also from the improved previous prediction, which facilitates the procedure. In particular, the matrix $\mathbf{W}_i(\mathbf{y}, \mathbf{x}_{i-1})$ and vector $\mathbf{w}_i(\mathbf{y}, \mathbf{x}_{i-1})$ are regressed at the i-th level of the cascade. Figure 7.4(a) depicts a schematic illustration of the non-blind RTF cascade model from Schmidt et al. (2013).



(a)



(b)

Figure 7.4: Comparison of standard, (Schmidt et al., 2013), versus our proposed interleaved RTF cascade schemes. (a) Standard non-blind RTF cascade: The blur kernel is set to the initial blur estimate $\mathbf{k}_0$ and stays invariant over the cascade. (b) Interleaved RTF cascade: The kernel is re-estimated over cascade stages using the refined image predictions $\mathbf{x}_i$. In the experiments, we mostly use Xu and Jia (2010) to obtain the initial blur estimate, but we can also initialize with the delta kernel by using several interleaved RTF cascades to operate over the scales of an image pyramid.

7.4.2 *Interleaved RTF Cascades*

Besides their quantitative and qualitative benefits in terms of image quality and efficiency, a distinctive feature of discriminative deconvolution methods is their adaptability to kernel estimation (errors). In particular, RTF cascades yield best results when trained with blur kernels of similar kind as those provided at test time (Schmidt et al., 2013). In this chapter, we *interleave* the image prediction steps in the RTF cascade with kernel re-estimation. Note that we here focus on uniform blur. Although the image formation model of spatially varying blur is more involved, the procedure detailed below is in principle equally valid.

We design an interleaved procedure by updating the blur kernel using the improved latent image prediction available at every cascade

level, such that each RTF stage is provided with a refined kernel estimate. Note that the cascade is initialized with the kernel estimate $\mathbf{k}_0$ of an auxiliary method[2]. For higher stages $i = 1, \dots, N$ of the interleaved cascade, the output $\mathbf{x}_i$ of the i-th RTF is used to compute a refined kernel estimate $\mathbf{k}_i$. Figure 7.4(b) depicts a schematic illustration of the interleaved RTF prediction cascade.

Specifically, we compute the kernel update using the image derivatives by minimizing with respect to $\mathbf{k}$ the objective function

$$f(\mathbf{k}) = \|\nabla \mathbf{y} - \mathbf{k} \otimes \nabla \mathbf{x}_i\|^2 + \gamma \|\mathbf{k}\|_1. \tag{7.4}$$

Hereby, we let $\nabla \mathbf{x} = (\partial_1 \mathbf{x}, \partial_2 \mathbf{x}) = (\mathbf{f}_1 \otimes \mathbf{x}, \mathbf{f}_2 \otimes \mathbf{x})$ denote the canonical image gradients computed with the standard derivative filters $\mathbf{f}_1 = [1, -1]$ and $\mathbf{f}_2 = [1, -1]^T$. For the gradient image, we define convolution by the blur kernel to apply component-wise, such that $\mathbf{k} \otimes \nabla \mathbf{x} = (\mathbf{k} \otimes \partial_1 \mathbf{x}, \mathbf{k} \otimes \partial_2 \mathbf{x})$. Note further that the objective for kernel re-estimation (Equation 7.4) consists of a squared residuals term motivated by a Gaussian noise assumption, and an L^1-norm penalty to encourage kernel sparsity, which is weighted by a constant $\gamma > 0$. The regularization parameter γ can be learned from data (see Section 7.4.3), but even simply setting $\gamma = 1$ already yields very good results. Algorithm 1 summarizes the proposed interleaved RTF cascade for blind deblurring.

We optimize the kernel update objective $f(\mathbf{k})$ of Equation (7.4) using iteratively reweighted least squares (IRLS). This means iteratively solving varying least-squares problems until the distance between consecutive solutions passes below a convergence threshold. Specifically, at the j-th iteration of IRLS, we compute

$$\mathbf{k}^j = \operatorname{argmin}_{\mathbf{k}} \|\nabla \mathbf{y} - \mathbf{k} \otimes \nabla \mathbf{x}_i\|^2 + \gamma \mathbf{k}^T \operatorname{diag}(\mathbf{z}) \mathbf{k}. \tag{7.5}$$

Thereby, the n-th element of the weighting vector $\mathbf{z}$ is determined by $z_n = 1 / \max(k_n^{j-1}, \epsilon)$, where we fixed $\epsilon = 10^{-5}$ in the experiments. Minimizing the quadratic expression in Equation (7.5) is equivalent to solving a linear system of equations $\mathbf{A}\mathbf{k} = \mathbf{b}$. The left-hand side matrix of this system is $\mathbf{A} = \sum_h \sum_c [\partial_h \mathbf{x}]_c [\partial_h \mathbf{x}]_c^T + \gamma \operatorname{diag}(\mathbf{z})$, using $[\cdot]_c$ to denote the c-th kernel-sized clique. On the other hand, the right-hand side vector is $\mathbf{b} = \sum_h \sum_c [\partial_h \mathbf{x}]_c (\partial_h \mathbf{y})_c$, using $(\partial_h \mathbf{y})_c$ to denote the pixel situated at the center of the c-th kernel-sized clique in the derivative image $\partial_h \mathbf{y}$. Note that this system is generally not amenable to solving by fast Fourier transform (FFT). Hence we use conjugate gradients with Jacobi preconditioning, computing the diagonal of the system matrix as $\sum_h \mathbf{1} \otimes [\partial_h \mathbf{x}]^{\circ 2} + \gamma \mathbf{z}$, where $[\cdot]^{\circ 2}$ denotes component-wise Hadamard square, while $\mathbf{1}$ constitutes an image of ones and has size $\dim(\mathbf{x}) - \dim(\mathbf{k}) + [1, 1]^T$.

2 In the experiments, we mostly use Xu and Jia (2010) to perform this step, but we can also initialize with the delta kernel when estimating the blur over the scales of an image pyramid. (In this case, the final kernel estimate at one scale is upsampled to serve as the initial estimate for the next interleaved RTF cascade, see Figure 7.9.)

Algorithm 1 Interleaved RTF cascade (Test-time deblurring)

input: Blurry image $\mathbf{y}$, initial blur kernel $\mathbf{k}_0$
output: Deblurred image $\mathbf{x}_N$, refined blur kernel $\mathbf{k}_N$

for $i = 1, ..., N$ **do**
[update latent image using i-th RTF regressor]

$$\mathbf{x}_i := \left(\mathbf{W}_i(\mathbf{y}, \mathbf{x}_{i-1}) + \frac{1}{\sigma^2} \mathbf{T}_{\mathbf{k}_{i-1}}^\top \mathbf{T}_{\mathbf{k}_{i-1}} \right)^{-1} \times \left(\mathbf{w}_i(\mathbf{y}, \mathbf{x}_{i-1}) + \frac{1}{\sigma^2} \mathbf{T}_{\mathbf{k}_{i-1}}^\top \mathbf{y} \right)$$

[update blur kernel]

$$\mathbf{k}_i := \operatorname{argmin}_{\mathbf{k}} \|\nabla \mathbf{y} - \mathbf{k} \otimes \nabla \mathbf{x}_i\|^2 + \gamma \|\mathbf{k}\|_1$$

end for

7.4.3 *Learning*

TRAINING DATA. We compiled sharp images for use as ground-truth data from two different benchmark datasets, the Berkeley segmentation data set and benchmark (BSDS 500, Arbelaez et al., 2011), and PASCAL visual object classes (PASCAL VOC, Everingham et al., 2010, 2015). Please note that the training images stem from entirely different sources than those used in the experimental evaluation (Section 7.5). As blur data we used 95 realistic blur kernels generated by recording the trajectory of a light source under human camera shake (see Section 7.3). We complemented these with synthetic blurs created by projecting randomly sampled motions in 3D space onto the camera plane (Schmidt et al., 2013). Note that none of these kernels is used at test time. To obtain blurry images, we synthetically convolved the ground-truth images and added Gaussian noise of standard deviation equal to 0.2% of the maximum pixel intensity. We used 336 clean and corrupted image pairs and blur kernels to train our models.

Table 7.1: Average PSNR (dB) values on the test set of Levin et al. (2011).

Method	∅ PSNR
Fergus et al. (2006)	29.38
Cho and Lee (2009)	29.71
Xu et al. (2013)	29.74
Levin et al. (2011)	30.05

Method	∅ PSNR
Standard RTF	31.16
Interleaved RTF	**31.50**

LEARNING THE LATENT IMAGE PREDICTION. At each level of the cascade, we learn a separate RTF model for image restoration. Besides the blurry input image, each model receives as additional input

Table 7.2: Average PSNR (dB) values on the test set of Köhler et al. (2012).

Method	∅ PSNR
Krishnan et al. (2011)	25.73
Hirsch et al. (2011)	27.77
Whyte et al. (2011)	28.07
Cho and Lee (2009)	28.98
Xu and Jia (2010)	29.54

Method	∅ PSNR
Std. RTF	29.91
Interlvd. RTF	**30.11**

the previous image prediction and is further parameterized by a blur kernel of increasing refinement. This is different from Schmidt et al. (2013), where the blur kernel remains fixed throughout all stages. We remark that the RTF models learned at every level adapt precisely to the kernel re-estimation and to the preceding image predictions given as inputs (see Table 7.3). The resulting, interleaved cascade thus forms a unit of inter-related components and needs to be trained together. We opt for regression trees of depth 7. To leverage more discriminative features than simple pixel intensities, we rely on the Field of Experts (FOE) filter bank of Gao and Roth (2012), *i.e.*, each model receives as additional features the filter responses of the previous prediction. Per depth level, we use 40 iterations of the limited-memory Broyden–Fletcher–Goldfarb–Shanno (LBFGS) algorithm to optimize the model parameters, with another 100 clean-up cycles after splitting the leaves at the final level 7. To accelerate the learning procedure, we did not use the original size images, but 125×125 sized pairs of degraded and sharp crops. Learning a cascade of depth 3 (plus evaluating the full interleaved model on the training images for each additional level) took 10 days on a machine with a 3.20GHz Intel Core i7 3930K CPU. Training time could be reduced by parallel computing on several machines.

LEARNING THE BLUR KERNEL UPDATE. With regard to updating the blur kernel, the regularization parameter γ weighting the influence of likelihood and prior in the objective function for the kernel update (Equation 7.4) may also be learned from data in a loss-based fashion. We opt for a blur kernel loss function based on the outlier resistant L1 metric, namely $\epsilon(\mathbf{k}, \mathbf{k}_{gt}) = \|\mathbf{k} - \mathbf{k}_{gt}\|_1 / |\mathbf{k}|$, where $|\mathbf{k}|$ denotes the number of kernel elements. Note here that care must be taken to align the blur kernels with each other before evaluating the distance, since a translation in the kernel simply leads to a translation in the deblurred image, and this should not be penalized. Although we could also optimize w.r.t. image quality, it is more efficient to compute the kernel loss, which obviates the more expensive image prediction step.

Hence at the i-th level of the interleaved cascade, a weight γ_i can be learned to optimize the empirical risk $\langle \epsilon(\mathbf{k}_i, \mathbf{k}_{gt}) \rangle = \frac{1}{N} \sum_n \epsilon(\mathbf{k}_i^n, \mathbf{k}_{gt}^n)$ over the training data. Since this is a unimodal objective function, a simple 1D line search suffices to find the optimum.

Table 7.3: Performance of RTF cascade models in average PSNR (dB) on the test set of Levin et al. (2011). Training and prediction are performed with interleaved or without re-estimation of the blur kernel over cascade levels.

	Prediction	
Training	Interleaved	Standard
Interleaved	**31.50**	30.67
Standard	30.81	**31.16**

7.5 EXPERIMENTS

Table 7.1 gives the performance of the proposed, interleaved RTF cascade on the benchmark of Levin et al. (2011). Our interleaved algorithm outperforms the blind deblurring methods Fergus et al. (2006); Levin et al. (2011); Xu et al. (2013); Cho and Lee (2009) on this benchmark with a very large margin of at least 1.45 dB. We further evaluated the non-blind, standard RTF cascade on this benchmark, using the blur estimate of Xu and Jia (2010) as input. This guarantees a fair comparison to the interleaved RTF cascade, which, although bootstrapped with Xu and Jia (2010), re-estimates the blur iteratively over the prediction stages. We remark that standard RTF cascades are state-of-the-art in non-blind deblurring and outperform many existing sparsity-based methods (Schmidt et al., 2013). Table 7.1 shows that our interleaved RTF cascade achieves significantly better results than the state-of-the-art non-blind cascade of Schmidt et al. (2013) by 0.34 dB in PSNR. This demonstrates how useful it is to re-estimate the blur kernel between discriminative image updates in a learned cascade.

Table 7.2 gives results on the benchmark of Köhler et al. (2012). Here, our interleaved algorithm achieves substantially better results than a multitude of other methods (Xu and Jia, 2010; Krishnan et al., 2011; Hirsch et al., 2011; Cho and Lee, 2009; Whyte et al., 2011) by at least 0.57 dB. Note that several images of the dataset of Köhler et al. (2012) are very challenging, having spatially varying blur of over 100 pixels. The interleaved algorithm again outperforms its standard, non-blind counterpart by a significant margin of 0.2 dB.

Figures 7.1, 7.5 and 7.6 show that our method preserves challenging regions of image texture faithfully, while suppressing ringing and noise artifacts in smooth regions or on the image boundary. Notably,

(a) Sharp reference (b) Blurry input

(c) Interleaved RTF (d) Xu and Jia (2010)

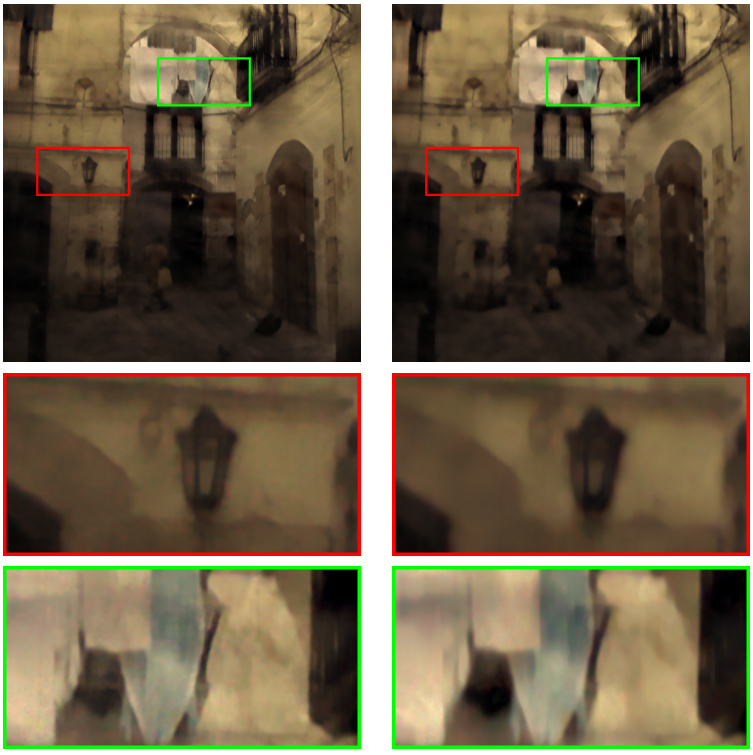
(e) Cho and Lee (2009) (f) Whyte et al. (2011)

(g) Hirsch et al. (2011) (h) Krishnan et al. (2011)

Figure 7.5: Qualitative comparison of deconvolution algorithms on a benchmark image of Köhler et al. (2012). The reference image shown in (a) is the first frame of the recorded motion. The interleaved RTF simultaneously recovers sharp edges such as the patterns of the flags, while keeping boundary artifacts at a minimum.

(a) Sharp reference

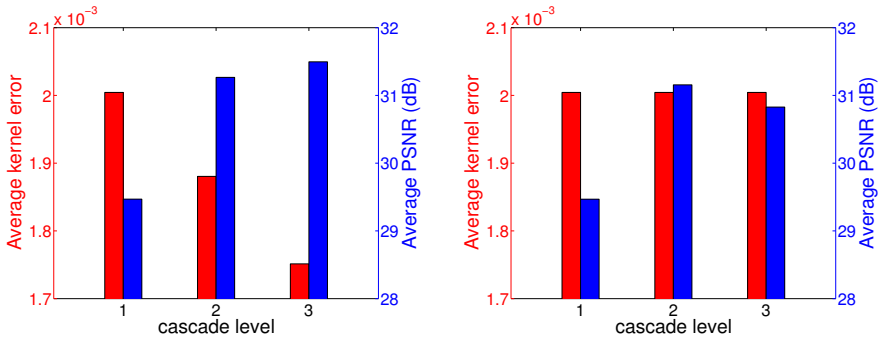
(b) Blurry



(c) Interleaved

(d) Standard

Figure 7.6: Qualitative comparison of interleaved versus standard RTF cascade on a benchmark image (Köhler et al., 2012). The interleaved RTF cascade recovers a higher level of image details and yields a more realistic deblurring result.



(a) Interleaved RTF cascade (b) Standard RTF cascade

Figure 7.7: Average blur kernel error versus image quality over interleaved and standard RTF cascade levels on the test set of Levin et al. (2011). The kernel error is quantified in mean absolute distance $\epsilon(\mathbf{k}, \mathbf{k}_{gt}) = \|\mathbf{k} - \mathbf{k}_{gt}\|_1 / |\mathbf{k}|$ to the ground truth blur (letting $|\mathbf{k}|$ denote the number of kernel elements). The interleaved RTF cascade simultaneously enhances the image and blur kernel.

Figure 7.5 demonstrates visibly superior performance of the interleaved cascade over a wide variety of blind deconvolution methods, while Figure 7.6 shows that interleaving with kernel updates yields a noticeably higher degree of realism in the deblurred image than using the standard cascade.

We further analyze the benefit of custom, discriminative training of the interleaved cascade to the refined kernel estimates available at each stage. Table 7.3 gives results for prediction with and without interleaved kernel updates, using RTF cascades learned with and without interleaved kernel updates. We observe that it is important to train the image prediction step based on the refined blur estimates to unlock the full potential of our approach. Simply interleaving a pre-trained standard cascade with blur updates leads to substantially inferior results. Note further that learning the image restoration steps expressly to extract maximum effect from the refined kernel estimates is a key benefit of discriminative updates.

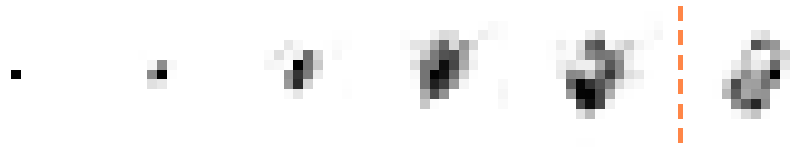
To gain more insight into the role of kernel refinement over cascade stages, we rely on the dataset of Levin et al. (2011), since it includes ground truth blur kernels to evaluate with. In particular, we measure the mean absolute distance of the (aligned) blur estimates to the ground truth kernels. Figure 7.7 depicts the average kernel error versus the average image quality over all 32 image and kernel pairs of the benchmark, shown after each of three cascade levels. We observe that the increasing image quality over the cascade allows to improve the kernel estimate and vice versa, while on the other hand, holding the blur fixed over the cascade leads to inferior overall performance.

We further examine the blur refinement effect of our algorithm in a visual study, relying once more on the 8 ground-truth camera shakes

(1) Xu & Jia [30] (2) proposed (3) ground truth

Figure 7.8: Kernel refinement on the dataset of Levin et al. (2009). For all of the 8 blurs in the test set, a triple is displayed horizontally. From left to right, each triple consists of: (1) The estimate of Xu and Jia (2010) used to initialize the interleaved cascade, (2) the refined blur estimate at the final level of the cascade, (3) the ground-truth kernel. Each triple is scaled jointly to the full intensity range.

(a) Blurry, SSIM 0.841 (b) Deblurred, SSIM 0.943



(c) Kernel estimates over scale. Rightmost: Ground-truth blur.

Figure 7.9: Multiscale interleaved RTF regression with *delta kernel initialization*. Each level of the pyramid is equipped with a progressively more powerful interleaved RTF cascade. Image and ground-truth kernel from Levin et al. (2009).

of Levin et al. (2011). Fig 7.8 shows three versions of each camera movement: The kernel estimate of Xu and Jia (2010) used as initialization to the interleaved restoration process, the blur estimate from the last stage of the interleaved cascade, and the ground-truth kernel provided with the benchmark. We observe that the interleaved estimation procedure substantially enhances the initial estimate.

To measure running times, we used a 3.20GHz Intel Core i7 3930K processor. For a kernel size of 41×41, blind deconvolution with our interleaved cascade algorithm needed 98.66s for an image of size 800×800. For comparison, we measured 156.49s for the efficient deblurring algorithm of Krishnan et al. (2011). Note that as a prototype, our implementation is not optimized for fast running time.

Finally, to demonstrate that our approach does not require a specific auxiliary method for initialization, Figure 7.9 shows an instance of multiscale interleaved RTF regression with delta kernel initialization. Hereby, interleaved cascades are used to predict image and kernel estimates at each level of a Gaussian pyramid. The estimates at one level are enlarged to serve as inputs for the next finer level. Note that the model trained with initial blur estimation from another method (Xu and Jia, 2010) cannot be used with delta blur initialization. Instead, going from coarse to fine, we trained progressively more powerful interleaved models to account for the higher level of image details and texture.

7.6 CONCLUSION

In this chapter, we put forth a novel, interleaved RTF cascade model for blind deblurring that consolidates discriminative image prediction with blur estimation, whereby each step is trained expressly to fit to the other. The model is validated by extensive experimentation, namely *(1)* quantitative insights into the effects of interleaving with kernel updates, *(2)* solid results on two standard benchmarks, and *(3)* qualitative comparisons to other deblurring methods including standard RTF cascades. To train our model, we used a novel dataset of human camera shakes, which was generated by recording LED trajectories with a handheld camera. Code and data are available online[3].

Future work must first investigate predicting not only the image, but also the kernel by use of the RTF framework. This idea is given a more detailed treatment in a separate paragraph of Section 8.2. Another item of future research is to generalize the proposed interleaved framework to spatially-varying blur. In that regard, kernel data for discriminative training may be obtained by supplanting the single point light source by a grid of LEDs. Additionally, the uniform image formation model must be replaced by a spatially-varying one (*e.g.*, Hirsch et al., 2011).

3 https://bitbucket.org/visinf/projects-interleaved-rtf

8 CONCLUSION

CONTENTS

These final observations first of all provide a summary of the contributions made in this thesis. Thereby, we *(1)* place the content of each chapter into a larger context, *(2)* identify the overarching themes and connecting elements of this dissertation, and *(3)* make reference to relevant current developments in research and technology. The second component of this chapter provides a rough plan of potential future approaches to extend our research into novel directions beyond the results of this thesis. The ideas on further research presented below critically reexamine some of the material contained in earlier chapters. Throughout our review of previous results and while outlining directions of future research, we cite selected publications to clarify our contributions in relation to the research field. The coverage of related work is somewhat reduced in this closing text. Please refer to earlier chapters for more details on the relevant literature.

8.1 SUMMARY

8.1.1 *Foundations of Low-level Vision*

Our first contribution in Chapter 3 of this dissertation may be categorized as pure research into low-level vision. To be specific, we established a link between two model types which permeate the field, namely Markov random fields (MRFs) on the one hand, and variational models on the other.

CONTRIBUTIONS. Let us reiterate a critical difference, namely that variational models operate on functions defined *almost everywhere* on the continuous domain, whereas MRFs express the probabilistic dependencies of a discrete set of random variables. In Chapter 3, we connected the two classes by *(1)* approximating spatially continuous functions via linear combinations of finite elements with local area

of support, *(2)* equating the basis coefficients with pixel intensities, and *(3)* integrating over the domain variable. These measures lead us from a given variational model to its equivalent MRF energy. Our implementation of the derived connection placed primary emphasis on robust penalty functions. This goes beyond previous work on model connections focusing on quadratic regularizers (Szeliski, 1990a). In particular, we derived MRFs corresponding to the ubiquitous total variation (TV) model (Rudin et al., 1992) and drew samples to estimate first and second-order moments, thereby illustrating the access to probabilistic inference by virtue of the constructed connection. We further modified the duality based maximum a-posteriori (MAP) approach of Chambolle (2004) as a step toward fast inference for a variant of TV potential ensuing from finite element discretization.

8.1.2 *High-order Bayesian Restoration*

Having examined a connection between two basic models for low-level vision in Chapter 3, we continued with more specialized research into image restoration in Chapters 4 and 5, where we considered high-order, filter-based MRF models. Recall that the merit of expanding from conventional pairwise to larger filters is heightened model expressiveness, as simple gradient models tend to produce piece-wise constant restoration results, whereas wider ranging neighborhoods also account for more intricate image structures such as texture. Of course, such high-order models need to be correctly designed as a prerequisite. An established design approach is the Field of Experts (FOE) framework (Roth and Black, 2009), one aspect of which is that the resulting, generatively learned MRFs perform best under Bayesian estimation (Schmidt et al., 2010). In accordance with this insight, we presented an extensive study of two implementations of Bayesian inference in Chapters 4 and 5, namely *Gibbs sampling* and *mean field*, respectively.

GAUSSIAN SCALE MIXTURES. Focal to both procedures is the use of Gaussian scale mixtures (GSMs) (*e.g.*, Portilla et al., 2003), which has twofold reasons. First, this function class comprises sufficiently kurtotic elements to suitably reflect the statistics of natural images. Second, GSMs are amenable to expansion by latent variables indicating a mixture component. The correspondingly augmented joint distributions have the crucial property of reducing to Gaussians conditioned on the indicators. This enables both, drawing samples as part of the Gibbs process, and updating the approximate distribution as part of the mean field procedure. Let us proceed by reviewing our findings in more detail.

CONTRIBUTIONS. In Chapter 4, we made use of Gibbs sampling to carry out non-blind deblurring in a Bayesian setting with a high-order, generative FOE. We documented competitive results in comparison to a variety of standard MAP algorithms (*e.g.*, Levin et al., 2007; Krishnan and Fergus, 2009; Roth and Black, 2009). From a qualitative viewpoint, both smooth and textured image regions were faithfully reconstructed by our approach. However, it is no surprise that Gibbs sampling incurs high runtime requirements. To address this issue, we studied mean field as a Bayesian inference method in Chapter 5. Thereby, we measured a considerable runtime speed-up and ascertained that in an image restoration setting, the update operations have quadratic complexity in the clique size, which indicates graceful scaling to larger cliques. Experiments in denoising attested to better performance than employing MAP estimation for the same generative FOE, while remaining in competitive distance to the more thorough, but computationally intensive Gibbs sampling algorithm. We validated the mean field approach extensively in denoising, non-blind deblurring, noise estimation and layer separation, while conducting numerous comparisons to other methods in the literature (*e.g.*, Levin et al., 2007; Ishikawa, 2009; Zoran and Weiss, 2009).

8.1.3 *Blind Image Deconvolution*

Whereas the contributions advanced in Chapters 3 through 5 essentially revolve around applications of non-blind nature such as denoising or deblurring with known kernel[1], we subsequently turned our focus to blind image deconvolution. In that regard, one promising avenue of research is to mitigate the problem complexity by augmenting the traditional camera hardware, *e.g.*, with on-board motion sensors, dual-lens technology, or a coded aperture. We remark that many such hardware measures rely on further processing, and it is reasonable to assume that in the future, camera technology will evolve to become heavily interwoven by software and computation. On the latter subject, we presented two purely algorithmic procedures for image deblurring in Chapters 6 and 7, which are of Bayesian and discriminative design, respectively. The two methods also address different respective blur categories, and we review them in more detail below.

LOCALIZED BLUR REMOVAL. In Chapter 6, we considered blur generated from fast moving objects in the scene, or from camera defocus occurring for regions of interest situated outside the depth of field. We referred to this type of blur as *localized*, since it is restricted to certain image area. For such image degradation, any successful restoration approach must both identify the blurred pixels and compute a

1 Recall that denoising can be viewed as the degenerate case of non-blind deblurring with the identity kernel.

corresponding kernel estimate. To achieve this, we proposed a new generative model equipped with a set of latent indicator variables to designate pixel-wise which blur is active. In contrast to various other methods relying on further data, such as user input (Dai and Wu, 2009), or specialized hardware (Martinello and Favaro, 2011), our approach only reads in the pixels as input. Another distinguishing factor, which produced a performance advantage in the experimental comparison, is that our method estimates non-parametric, freely varying kernels, whereas other image-based procedures tend to assume a fixed family of blurs, *e.g.*, box filters stemming from a finite, discrete range of speeds (Chakrabarti et al., 2010; Couzinié-Devy et al., 2013). Our Bayesian approach proved to be versatile in the sense that it can handle both motion and defocus blur concurrently.

DISCRIMINATIVE BLIND DEBLURRING. While Bayesian methods generalize well and allow to create synthetic data, their downside is often the cost in computation time. Application-specific discriminative models, on the other hand, frequently have favorable performance in both image quality and runtime. Hence in Chapter 7, we investigated a discriminative approach to blind removal of camera shake. We opted for the Regression Tree Field (RTF) framework (Jancsary et al., 2012a) to realize the required prediction steps, since RTFs combine the accuracy of regression trees with the tractability of Gaussian conditional random fields (CRFs). Our main contribution is to generalize the stacked RTF cascade architecture (Schmidt et al., 2016), which was previously used exclusively for non-blind deblurring, to the blind scenario by interleaving image predictions with kernel re-estimation. This is one of the first discriminative approaches to blind deblurring (see also Zuo et al. (2015); Xiao et al. (2016); Schuler et al. (2016)).

8.2 PERSPECTIVES AND FUTURE WORK

INCLUSIVE KULLBACK-LEIBLER DIVERGENCE. The Bayesian techniques studied in Chapters 4 and 5 require estimating expected values. Similarly, the kernel estimation procedure of Chapter 6 involves integrating approximately over image variables as suggested by the framework of Levin et al. (2009). Throughout this thesis, we have relied on sampling-based inference or mean field to calculate such approximations. The latter algorithm minimizes the so-called *exclusive* Kullback-Leibler (KL) divergence $\mathrm{KL}(q\|p)$ between the approximate and true distributions q and p, respectively. However, it can be argued (Minka, 2005) that for estimating moments or marginals, a more appropriate optimization target is $\mathrm{KL}(p\|q)$, which differs in a deceptively simple swap of the variables. This measure is termed *inclusive* KL-divergence, since it leads to proxy distributions extending

across several modes of the true probability. For the case of models with continuous variables, which have been our focus in this thesis, a procedure to minimize the inclusive distance term is expectation propagation (EP), initially conceived as a message-passing scheme (Minka, 2001). We note that the double loop variant advanced by Seeger and Nickisch (2011) is fast and guaranteed to converge, while Papandreou and Yuille (2011) further demonstrate how to robustly perform the required variance computation. As both of the latter publications consider log-concave potentials, particularly Laplacians, an unanswered question is whether EP is an effective method for the *learned* FOE potentials considered in this dissertation, which are generally less controllable and tougher to optimize. Future work must also investigate in how far EP leads to refined kernel estimates when utilized in a marginalized MAP approach as followed in Chapter 6. A positive answer would have a broad impact on the study of deconvolution problems ranging from reflectometry (Romeiro and Zickler, 2010) to image restoration (Wipf and Zhang, 2014).

REGRESSION TREE FIELD KERNEL PREDICTION. One intriguing aspect of the interleaved cascade discussed in Chapter 7 is how the image predictions attenuate errors in the kernel estimation. Although we have already trained the regularization parameter in a loss-based fashion (see Section 7.4.3), a sound question is whether we can do better by accomplishing the kernel update directly within the RTF framework. To that end, we remark that the deblurring likelihood is proportional to a Gaussian in the kernel $\mathbf{k}$,

$$p(\mathbf{y}|\mathbf{x},\mathbf{k}) \propto \mathcal{N}(\mathbf{y};\mathbf{k}\otimes\mathbf{x},\sigma^2\mathbf{I}) \propto \mathcal{N}(\mathbf{k};\mathbf{M}^{-1}\mathbf{v},\mathbf{M}^{-1}). \tag{8.1}$$

Thereby, we used the notation $\mathbf{M} = \frac{1}{\sigma^2}\sum_i \mathbf{x}_i\mathbf{x}_i^\mathsf{T}$ and $\mathbf{v} = \frac{1}{\sigma^2}\sum_i \mathbf{x}_i y_i$, where $\mathbf{x}_i$ denotes the i-th kernel-sized hidden image clique transformed to a column vector, and y_i is the pixel at the center of the corresponding clique in the blurry image $\mathbf{y}$. In the following, let us assume that we have at our disposal earlier estimates $\mathbf{k}_0$ and $\mathbf{x}$ of the kernel and latent image, respectively[2]. To obtain the blur update, we may appeal to a generative approach, similar to (Schmidt et al., 2016). Assuming a Gaussian kernel prior $p(\mathbf{k}) = \mathcal{N}(\mathbf{k};\mathbf{W}^{-1}\mathbf{w},\mathbf{W}^{-1})$, we verify that

$$\begin{aligned} p(\mathbf{k}|\mathbf{y},\mathbf{x}) &\propto p(\mathbf{y}|\mathbf{x},\mathbf{k})p(\mathbf{k}) && (8.2)\\ &\propto \mathcal{N}(\mathbf{k};\mathbf{M}^{-1}\mathbf{v},\mathbf{M}^{-1})\cdot\mathcal{N}(\mathbf{k};\mathbf{W}^{-1}\mathbf{w},\mathbf{W}^{-1}) && (8.3)\\ &\propto \mathcal{N}(\mathbf{k};(\mathbf{M}+\mathbf{W})^{-1}(\mathbf{v}+\mathbf{w}),(\mathbf{M}+\mathbf{W})^{-1}). && (8.4)\end{aligned}$$

At this point, the matrix $\mathbf{W} = \mathbf{W}(\mathbf{k}_0)$ and vector $\mathbf{w} = \mathbf{w}(\mathbf{k}_0)$ can be regressed from the earlier kernel estimate as CRF parameters using

2 For process initialization, the blur and sharp image can be set to the identity kernel and input image *resp.*, *i.e.*, $\mathbf{k}_0 \equiv \delta$, and $\mathbf{x} \equiv \mathbf{y}$. Kernel estimation can be pursued over multiple scales, as is standard.

the RTF framework as desired initially. Preliminary experiments suggest the use of mean absolute error for loss-specific learning, in case a distance measure to ground-truth kernels is to be optimized. Another option as target function is a quality metric operating on the restored images directly, such as PSNR or SSIM; this is likely more challenging with regard to training. Note that previous work on kernel fusion with RTFs (Mai and Liu, 2015) differs from the procedure outlined above in relying on the outcomes of several deblurring algorithms running in advance. A further distinction is that our design connects the current image estimate with the blur prediction step. If the proposed approach succeeds in recovering kernels of improved clarity, one can expect substantial performance gains.

USER-SPECIFIC CAMERA MOTION. A well-established fact in experimental psychology is that humans can recognize familiar people based on their gait, that is, by seeing only the motion of point light sources attached to an individual's limbs (Johansson, 1973). Computer vision systems are also capable of person recognition from stride (Wang et al., 2003; Man and Bhanu, 2006; Tao et al., 2007). However, to the best of our knowledge, it is an open question whether camera shake ultimately has a comparable, biometric nature, although a preliminary analysis hints at individual traits (Horstmeyer, 2010). If corroborated by a larger study, one could investigate the viability of a deblurring system learning over time to compensate optimally for user-specific hand motion during exposure time. This would constitute a step toward building a personalized camera with the aid of machine learning.

A

CHAMBOLLE'S ALGORITHM EXTENDED TO FINITE ELEMENTS

Chambolle's duality-based minimization algorithm for total variation (TV) (Chambolle, 2004) can be adapted for maximum a-posteriori (MAP) inference in linear FE-MRFs. Although Subsection 3.4.1 provides an overview on how to extend the original technique, for the convenience of the reader, we here further illustrate the arguments. (For the necessary concepts from convex analysis, such as conjugates and subdifferentials, see *e.g.* Hiriart-Urruty and Lemaréchal (1993a,b)). Computing the MAP estimate of the Markov random field (MRF) obtained from the Rudin-Osher-Fatemi (ROF) model (Rudin et al., 1992) consists of solving

$$\min_{\mathbf{o}\in\mathbf{X}} \frac{1}{2\lambda}(\mathbf{o}-\mathbf{i})^{\mathrm{T}}\mathbf{A}(\mathbf{o}-\mathbf{i}) + E_S(\mathbf{o}), \tag{A.1}$$

where $E_S(\mathbf{o})$ is the spatial term from Equation (3.26) with $\varphi(\mathbf{y}) = |\mathbf{y}|$. Taking note that $\mathbf{o}, \mathbf{i}$ are matrices in $\mathbf{X} = \mathbb{R}^{N\times N}$, and following the design of Chambolle (2004), we define a gradient operator

$$\nabla : \mathbf{X} \to \mathbf{Y}, \text{ where } \mathbf{Y} = \mathbb{R}^{N\times 2N} \times \mathbb{R}^{N\times 2N}, \tag{A.2}$$

that computes both forward and backward differences as

$$(\nabla\mathbf{o})_{i,j} = \frac{1}{2}\left((\nabla\mathbf{o})^1_{i,j}, (\nabla\mathbf{o})^2_{i,j}\right). \tag{A.3}$$

Here,

$$(\nabla\mathbf{o})^1_{i,j} = \begin{cases} o_{i+1,j} - o_{i,j} & \text{if } i,j < N, \\ o_{i,j-N} - o_{i-1,j-N} & \text{if } i > 1, j > N+1, \\ 0 & \text{otherwise,} \end{cases} \tag{A.4}$$

$$(\nabla\mathbf{o})^2_{i,j} = \begin{cases} o_{i,j+1} - o_{i,j} & \text{if } i,j < N, \\ o_{i,j-N} - o_{i,j-N-1} & \text{if } i > 1, j > N+1, \\ 0 & \text{otherwise.} \end{cases} \tag{A.5}$$

As a consequence, we obtain the formula

$$E_S(\mathbf{o}) = \sum_{i,j} \left|(\nabla\mathbf{o})_{i,j}\right|, \tag{A.6}$$

where $|\mathbf{y}| = \sqrt{y_1^2 + y_2^2}$ for all $\mathbf{y} = (y_1, y_2) \in \mathbb{R}^2$. The discrete divergence $\operatorname{div} : \mathbf{Y} \to \mathbf{X}$ is defined as the negative adjoint of ∇, which yields

$$(2\operatorname{div}(\mathbf{p}))_{i,j} \tag{A.7}$$

$$= \begin{cases} p^1_{i,j} - p^1_{i-1,j} & \text{if } 1 < i < N, j < N, \\ p^1_{i,j} & \text{if } i = 1, j < N, \\ -p^1_{i-1,j} & \text{if } i = N, j < N, \\ 0 & \text{otherwise} \end{cases} \tag{A.8}$$

$$+ \begin{cases} p^1_{i+1,j} - p^1_{i,j} & \text{if } 1 < i < N, N+1 < j, \\ p^1_{i+1,j} & \text{if } i = 1, N+1 < j, \\ -p^1_{i,j} & \text{if } i = N, N+1 < j, \\ 0 & \text{otherwise} \end{cases} \tag{A.9}$$

$$+ \begin{cases} p^2_{i,j} - p^2_{i,j-1} & \text{if } i < N, 1 < j < N, \\ p^2_{i,j} & \text{if } i < N, j = 1, \\ -p^2_{i,j-1} & \text{if } i < N, j = N, \\ 0 & \text{otherwise} \end{cases} \tag{A.10}$$

$$+ \begin{cases} p^2_{i,j+1} - p^2_{i,j} & \text{if } 1 < i, N+1 < j < 2N, \\ p^2_{i,j+1} & \text{if } 1 < i, j = N+1, \\ -p^2_{i,j} & \text{if } 1 < i, j = 2N, \\ 0 & \text{otherwise,} \end{cases} \tag{A.11}$$

for every $\mathbf{p} = (\mathbf{p}^1, \mathbf{p}^2) \in \mathbf{Y}$. In analogy to Chambolle (2004), we derive for the conjugate of E_S,

$$E_S^*(\mathbf{v}) = \begin{cases} 0 & \text{if } \mathbf{v} \in K, \\ +\infty & \text{otherwise}, \end{cases} \tag{A.12}$$

where $K = \left\{\operatorname{div}(\mathbf{p}) : \mathbf{p} \in \mathbf{Y}, \left|\mathbf{p}_{i,j}\right| \leqslant 1\right\}$. A sufficient condition for a minimum of Equation (A.1) is

$$0 \in \mathbf{A}(\mathbf{o} - \mathbf{i}) + \lambda \partial E_S(\mathbf{o}), \tag{A.13}$$

where 0 refers to the zero matrix. Equation (A.13) corresponds to

$$\frac{1}{\lambda}\mathbf{A}(\mathbf{i} - \mathbf{o}) \in \partial E_S(\mathbf{o}), \tag{A.14}$$

which is equivalent to

$$\mathbf{o} \in \partial E_S^*\left(\frac{1}{\lambda}\mathbf{A}(\mathbf{i} - \mathbf{o})\right). \tag{A.15}$$

This can be rewritten as

$$0 \in \frac{\mathbf{i}-\mathbf{o}}{\lambda} - \frac{\mathbf{i}}{\lambda} + \frac{1}{\lambda}\partial E_S^*\left(\frac{1}{\lambda}\mathbf{A}(\mathbf{i}-\mathbf{o})\right), \tag{A.16}$$

which is equivalent to

$$0 \in \mathbf{A}^{-1}\left(\frac{1}{\lambda}\mathbf{A}(\mathbf{i}-\mathbf{o})\right) - \frac{\mathbf{i}}{\lambda} + \frac{1}{\lambda}\partial E_S^*\left(\frac{1}{\lambda}\mathbf{A}(\mathbf{i}-\mathbf{o})\right). \tag{A.17}$$

If we define $\mathbf{v}$ as $\frac{1}{\lambda}\mathbf{A}(\mathbf{i}-\mathbf{o})$, the condition becomes

$$0 \in \mathbf{A}^{-1}\mathbf{v} - \frac{\mathbf{i}}{\lambda} + \frac{1}{\lambda}\partial E_S^*(\mathbf{v}), \tag{A.18}$$

which implies that $\mathbf{v}$ minimizes

$$\mathbf{v}^T\mathbf{A}^{-1}\mathbf{v} - \mathbf{v}^T\left(\frac{2\mathbf{i}}{\lambda}\right) + \frac{2}{\lambda}E_S^*(\mathbf{v}). \tag{A.19}$$

Once $\mathbf{v}$ is determined, the solution $\mathbf{o}$ follows by straightforward derivation. The properties of E_S^* yield that $\mathbf{v}$ is the element of K with minimal

$$Q(\mathbf{v}) := \mathbf{v}^T\mathbf{A}^{-1}\mathbf{v} - \mathbf{v}^T\left(\frac{2\mathbf{i}}{\lambda}\right). \tag{A.20}$$

Therefore we need to solve the constrained quadratic minimization problem

$$\min\left\{Q\left(\operatorname{div}(\mathbf{p})\right) : \mathbf{p} \in \mathbf{Y}, \left|\mathbf{p}_{i,j}\right|^2 - 1 \leqslant 0\right\}. \tag{A.21}$$

The Karush-Kuhn-Tucker conditions are

$$-\left(\nabla\left(\mathbf{A}^{-1}\operatorname{div}(\mathbf{p}) - \frac{\mathbf{i}}{\lambda}\right)\right)_{i,j} + \operatorname{diag}\left(\alpha_{i,j}\right)\mathbf{p}_{i,j} = 0, \tag{A.22}$$

where the non-negative $\alpha_{i,j}$ are Lagrange multipliers, and where we used that

$$\nabla_{\mathbf{p}}\left(Q\left(\operatorname{div}(\mathbf{p})\right)\right) = -\nabla\left(\left(\nabla_{\mathbf{p}}Q\right)\left(\operatorname{div}(\mathbf{p})\right)\right). \tag{A.23}$$

(Please note the distinction between the gradient w.r.t. $\mathbf{p}$ and the gradient operator defined in Equations (A.2)–(A.5)). A fixed point algorithm can now be derived just as in Chambolle (2004). We observe stable convergence for the same parameter $\tau = 0.25$ of the original paper.

B

MEAN FIELD UPDATES FOR LOCALIZED IMAGE BLUR REMOVAL

The objective of variational Bayesian (mean field) inference is to minimize the KL divergence between a tractable, approximate density q and the true distribution p. We choose a fully-factorized approximate density. Inference proceeds by updating groups of variables in turn, while keeping the others fixed. See Chapter 5 for more details on the inference procedure and how to derive the updates. We denote the approximate gradient and kernel distributions by $q(\partial_x) = \prod_j q(\partial_j \mathbf{x})$ and $q(\mathbf{k}) = \prod_i q(\mathbf{k}_i)$, where each of the factors is Gaussian with diagonal covariance, $q(\partial_j \mathbf{x}) = \mathcal{N}(\mathbf{n}_j, \mathbf{C}_j)$, and $q(\mathbf{k}_i) = \mathcal{N}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$. On the other hand, the indicator densities $q(\mathbf{h})$ and $q(\mathbf{t})$ are simply products of discrete distributions in each variable, $q(\mathbf{h}) = \prod_n q(\mathbf{h}_n)$ and $q(\mathbf{t}) = \prod_{n,j} q(\mathbf{v}_{nj})$.

B.1 BLUR INDICATORS

The update takes the form $q^*(\mathbf{h}_n) = \prod_i r_{ni}^{h_{ni}}$ with r_{ni} defined by

$$\begin{aligned} \log r_{ni} = &- \frac{1}{2\sigma^2} \sum_j \mathrm{E}_q \left[\left(\left(\partial_j y\right)_n - \mathbf{k}_i \otimes \left(\partial_j \mathbf{x}\right)_n \right)^2 \right] \\ &- \lambda \sum_{l:(l,n)\in\mathcal{N}} \sum_{\mathbf{h}_l} q(\mathbf{h}_l)[h_{li} \neq 1] + \text{const.} \end{aligned} \tag{B.1}$$

Thereby, the expectation E_q over all variables is

$$\begin{aligned} &\mathrm{E}_q \left[\left(\left(\partial_j y\right)_n - \mathbf{k}_i \otimes \left(\partial_j \mathbf{x}\right)_n \right)^2 \right] = \left(\mathbf{n}_{jn}^\mathsf{T} \boldsymbol{\mu}_i\right)^2 + \mathbf{n}_{jn}^\mathsf{T} \boldsymbol{\Sigma}_i \mathbf{n}_{jn} \\ &+ \boldsymbol{\mu}_i^\mathsf{T} \mathbf{C}_{jn} \boldsymbol{\mu}_i + \mathrm{Tr}(\mathbf{C}_{jn} \boldsymbol{\Sigma}_i) - 2(\partial_j y)_n \boldsymbol{\mu}_i^\mathsf{T} \mathbf{n}_{jn}. \end{aligned} \tag{B.2}$$

The n-th clique of $\mathbf{n}_j$ forms the column vector $\mathbf{n}_{jn}$, while the clique covariances form the diagonal matrix $\mathbf{C}_{jn}$.

B.2 BLUR KERNELS

To compute the update $q^*(\mathbf{k}_i) = \mathcal{N}\left(\boldsymbol{\mu}_i^*, \boldsymbol{\Sigma}_i^*\right)$, we use the auxiliary matrix and vector

$$\mathbf{A}_i = \sum_{n,j} r_{ni} \mathbf{n}_{jn} \mathbf{n}_{jn}^\mathsf{T} + r_{ni} \mathbf{C}_{jn} \tag{B.3}$$

$$\mathbf{b}_i = \sum_n r_{ni} \sum_j (\partial_j y)_n \mathbf{n}_{jn}. \tag{B.4}$$

The mean $\boldsymbol{\mu}_i^*$ is the solution to the quadratic program

$$\min \frac{1}{2}\boldsymbol{\mu}_i^\top \mathbf{A}_i \boldsymbol{\mu}_i - \mathbf{b}_i^\top \boldsymbol{\mu}_i \;\text{ subject to }\; \boldsymbol{\mu}_i \geqslant 0. \tag{B.5}$$

Further, $\operatorname{diag}\left(\boldsymbol{\Sigma}_i^*\right)$ is the component-wise inverse of $\operatorname{diag}\left(\mathbf{A}_i\right)$.

B.3 GSM INDICATORS

The update takes the form $q^*(\mathbf{v}_{nj}) = \prod_l \phi_{njl}^{\nu_{njl}}$, where

$$\phi_{njl} \propto \frac{\pi_l}{\sigma_l} \exp\left(-\frac{1}{2\sigma_l^2}\left(m_{jn}^2 + C_{jnn}\right)\right). \tag{B.6}$$

Here, C_{jnn} is the n-th diagonal entry of covariance $\mathbf{C}_j$.

B.4 GRADIENTS

To compute the update $q^*(\partial_j \mathbf{x}) = \mathcal{N}(\mathbf{n}_j^*, \mathbf{C}_j^*)$, we define the auxiliary matrix and vector

$$\mathbf{D}_j = \mathbf{M}_j + \frac{1}{\sigma^2}\sum_i \mathbf{T}_{\boldsymbol{\mu}_i}^\top \mathbf{R}_i \mathbf{T}_{\boldsymbol{\mu}_i} + \frac{1}{\sigma^2}\sum_i \boldsymbol{\Lambda}_i \tag{B.7}$$

$$\mathbf{e}_j = \frac{1}{\sigma^2}\sum_i \mathbf{T}_{\boldsymbol{\mu}_i}^\top \mathbf{R}_i \cdot \partial_j \mathbf{y}. \tag{B.8}$$

The n-th entry of the diagonal matrix $\mathbf{M}_j$ is $\sum_l \phi_{njl}/\sigma_l^2$, while additionally, $\mathbf{R}_i = \operatorname{diag}(r_{ni})$. The Toeplitz matrix $\mathbf{T}_{\boldsymbol{\mu}_i}$ denotes convolution by the kernel means $\boldsymbol{\mu}_i$. Further,

$$\boldsymbol{\Lambda}_i = \operatorname{diag}\left(\mathbf{T}_{\operatorname{diag}(\boldsymbol{\Sigma}_i)}^\top \operatorname{diag}\left(\mathbf{R}_i\right)\right), \tag{B.9}$$

where $\mathbf{T}_{\operatorname{diag}(\boldsymbol{\Sigma}_i)}$ denotes convolution by the kernel covariances. Then $\mathbf{n}_j^* = (\mathbf{D}_j)^{-1}\mathbf{e}_j$, and $\operatorname{diag}\left(\mathbf{C}_j^*\right)$ is determined as the component-wise inverse of $\operatorname{diag}\left(\mathbf{D}_j\right)$.

BIBLIOGRAPHY

Ansel Adams. *The Camera*, volume 1 of *The Ansel Adams Photography Series*. Little, Brown and Company, 1995a. ISBN 978-0821221846. (Cited on page 3.)

Ansel Adams. *The Negative*, volume 2 of *The Ansel Adams Photography Series*. Little, Brown and Company, 1995b. ISBN 978-0821221860. (Cited on page 3.)

Ansel Adams. *The Print*, volume 3 of *The Ansel Adams Photography Series*. Little, Brown and Company, 1995c. ISBN 978-0821221877. (Cited on page 3.)

Timo Ahonen, Esa Rahtu, Ville Ojansivu, and Janne Heikkilä. Recognition of blurred faces using local phase quantization. In *Proceedings of the International Conference on Pattern Recognition*, pages 1–4, Tampa, Florida, December 2008. doi: 10.1109/ICPR.2008.4761847. (Cited on page 2.)

Amazon.com, Inc. Annual report 2016, *Form 10-K*, April 2017. (Cited on page 1.)

Pablo Arbelaez, Michael Maire, Charless Fowlkes, and Jitendra Malik. Contour detection and hierarchical image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(5):898–916, May 2011. doi: 10.1109/TPAMI.2010.161. (Cited on pages x and 101.)

G. Aubert, A. El Hamidi, C. Ghannam, and M. Ménard. On a class of ill-posed minimization problems in image processing. *Journal of Mathematical Analysis and Applications*, 352(1):380–399, April 2009. doi: 10.1016/j.jmaa.2008.06.049. (Cited on pages 6, 33, 34, and 38.)

Gilles Aubert and Pierre Kornprobst. *Mathematical Problems in Image Processing: Partial Differential Equations and the Calculus of Variation*, volume 147 of *Applied Mathematical Sciences*. Springer New York, 2006. doi: 10.1007/978-0-387-44588-5. (Cited on page 6.)

S. Derin Babacan, Rafael Molina, Minh N. Do, and Aggelos K. Katsaggelos. Bayesian blind deconvolution with general sparse image priors. In A. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato, and C. Schmid, editors, *Proceedings of the 12th European Conference on Computer Vision*, volume 7577 of *Lecture Notes in Computer Science*, pages 341–355. Springer, 2012. doi: 10.1007/978-3-642-33783-3_25. (Cited on page 18.)

Saeideh Bakhshi, David A Shamma, Lyndon Kennedy, and Eric Gilbert. Why we filter our photos and how it impacts engagement. In *Proceedings of the Ninth International Conference on Web and Social Media*, pages 12–21, University of Oxford, Oxford, UK, May 2015. AAAI Press. (Cited on page 4.)

Leah Bar, Benjamin Berkels, Martin Rumpf, and Guillermo Sapiro. A variational framework for simultaneous motion estimation and restoration of motion–blurred video. In *Proceedings of the Eleventh IEEE International Conference on Computer Vision*, Rio de Janeiro, Brazil, October 2007. doi: 10.1109/ICCV.2007.4409009. (Cited on pages 23 and 80.)

David Barber. *Bayesian Reasoning and Machine Learning*. Cambridge University Press, 2012. ISBN 978-0521518147. (Cited on page 15.)

Benedicte Bascle, Andrew Blake, and Andrew Zisserman. Motion deblurring and super-resolution from an image sequence. In B. F. Buxton and R. Cipolla, editors, *Proceedings of the Fourth European Conference on Computer Vision*, volume 1065 of *Lecture Notes in Computer Science*, pages 571–582. Springer, 1996. doi: 10.1007/3-540-61123-1_171. (Cited on page 2.)

Klaus-Jürgen Bathe. *Finite Element Procedures*. Klaus-Jürgen Bathe, second edition, 2014. ISBN 978-0979004957. (Cited on pages 14, 15, 29, 30, and 34.)

Simon Beckouche, Jean-Luc Starck, and Jalal Fadili. Astronomical image denoising using dictionary learning. *Astronomy & Astrophysics*, 556(A132), August 2013. doi: 10.1051/0004-6361/201220752. (Cited on page 3.)

Anthony J. Bell and Terrence J. Sejnowski. An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7(6):1129–1159, November 1995. doi: 10.1162/neco.1995.7.6.1129. (Cited on page 19.)

M. Ben-Ezra and S.K. Nayar. Motion-based motion deblurring. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(6):689–698, June 2004. doi: 10.1109/TPAMI.2004.1. (Cited on page 49.)

Olivier Bernard, Denis Friboulet, Philippe Thévenaz, and Michael Unser. Variational B-spline level-set method for fast image segmentation. In *Proceedings of the 5th IEEE International Symposium on Biomedical Imaging*, pages 177–180, Paris, France, May 2008. doi: 10.1109/ISBI.2008.4540961. (Cited on page 30.)

Julian Besag. Spatial interaction and the statistical analysis of lattices. *Journal of the Royal Statistical Society. Series B, Statistical Methodology*, 36(2):192–236, 1974. URL http://www.jstor.org/stable/2984812. (Cited on pages 7, 16, 28, and 31.)

Michael J. Black and Anand Rangarajan. On the unification of line processes, outlier rejection, and robust statistics with applications in early vision. *International Journal of Computer Vision*, 19(1):57–91, July 1996. doi: 10.1007/BF00131148. (Cited on page 40.)

Andrew Blake, Pushmeet Kohli, and Carsten Rother, editors. *Markov Random Fields for Vision and Image Processing*. MIT Press, July 2011. ISBN 978-0262015776. (Cited on page 15.)

Yuri Boykov, Olga Veksler, and Ramin Zabih. Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(11):1222–1239, November 2001. doi: 10.1109/34.969114. (Cited on pages 15, 29, and 31.)

Kristian Bredies, Karl Kunisch, and Thomas Pock. Total generalized variation. *SIAM Journal on Imaging Sciences*, 3(3):492–526, September 2010. doi: 10.1137/090769521. (Cited on page 14.)

D. S. Broomhead and David Lowe. Radial basis functions, multivariable functional interpolation and adaptive networks. Memorandum 4148, Royal Signals and Radar Establishment, Malvern, Worcestershire, March 1988. (Cited on page 15.)

Tadhg Brosnan and Da-Wen Sun. Improving quality inspection of food products by computer vision—a review. *Journal of Food Engineering*, 61(1):3–16, January 2004. doi: 10.1016/S0260-8774(03)00183-3. (Cited on page 2.)

M. D. Buhmann. Radial basis functions. *Acta Numerica*, 9:1–38, January 2000. (Cited on page 15.)

Harold Christopher Burger, Bernhard Schölkopf, and Stefan Harmeling. Removing noise from astronomical images using a pixel-specific noise model. In *Proceedings of the IEEE International Conference on Computational Photography*, San Francisco, CA, April 2011. doi: 10.1109/ICCPHOT.2011.5753128. (Cited on page 3.)

Patrizio Campisi and Karen Egiazarian, editors. *Blind Image Deconvolution: Theory and Applications*. CRC Press, May 2007. ISBN 978-0849373671. (Cited on page 19.)

Ayan Chakrabarti, Todd Zickler, and William T. Freeman. Analyzing spatially-varying blur. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 2512–2519, San Francisco, California, June 2010. doi: 10.1109/CVPR.2010.5539954. (Cited on pages 6, 9, 23, 79, 80, 84, 87, 88, 89, and 114.)

Antonin Chambolle. An algorithm for total variation minimization and applications. *Journal of Mathematical Imaging and Vision*, 20

(1):89–97, January 2004. doi: 10.1023/B:JMIV.0000011325.36760.1e. (Cited on pages 6, 14, 29, 32, 33, 38, 42, 112, 117, 118, and 119.)

Antonin Chambolle and Pierre-Louis Lions. Image recovery via total variation minimization and related problems. *Numerische Mathematik*, 76(2):167–188, April 1997. doi: 10.1007/s002110050258. (Cited on page 14.)

R. H. Chan, T. F. Chan, and Chiu-Kwong Wong. Cosine transform based preconditioners for total variation deblurring. *IEEE Transactions on Image Processing*, 8(10):1472–1478, October 1999. doi: 10.1109/83.791976. (Cited on page 14.)

Tony Chan, Antonio Marquina, and Pep Mulet. High-order total variation-based image restoration. *SIAM Journal on Scientific Computing*, 22(2):503–516, 2000. doi: 10.1137/S1064827598344169. (Cited on page 14.)

Tony F. Chan, Selim Esedoglu, and Frederick Park. A fourth order dual method for staircase reduction in texture extraction and image restoration problems. In *Proceedings of the IEEE International Conference on Image Processing*, pages 4137–4140, Hong Kong, September 2010. doi: 10.1109/ICIP.2010.5653199. (Cited on page 14.)

Giannis Chantas, Nikolaos Galatsanos, Aristidis Likas, and Michael Saunders. Variational Bayesian image restoration based on a product of t-distributions image prior. *IEEE Transactions on Image Processing*, 17(10):1795–1805, October 2008. doi: 10.1109/TIP.2008.2002828. (Cited on page 65.)

Pierre Charbonnier, Laure Blanc-Féraud, Gilles Aubert, and Michel Barlaud. Deterministic edge-preserving regularization in computed imaging. *IEEE Transactions on Image Processing*, 6(2):298–311, February 1997. doi: 10.1109/83.551699. (Cited on page 38.)

Subhasis Chaudhuri, Rajbabu Velmurugan, and Renu Rameshan. *Blind Image Deconvolution: Methods and Convergence*. Springer International Publishing, 2014. doi: 10.1007/978-3-319-10485-0. (Cited on page 19.)

Sunghyun Cho and Seungyong Lee. Fast motion deblurring. *ACM Transactions on Graphics*, 28(5):145:1–145:8, December 2009. doi: 10.1145/1618452.1618491. (Cited on pages 21, 49, 80, 95, 101, 102, 103, and 104.)

Taeg Sang Cho, Neel Joshi, C. Lawrence Zitnick, Sing Bing Kang, Richard Szeliski, and William T. Freeman. A content-aware image prior. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 169–176, San Francisco, California, June 2010. doi: 10.1109/CVPR.2010.5540214. (Cited on pages 48 and 49.)

Taeg Sang Cho, Sylvain Paris, Berthold K. P. Horn, and William T. Freeman. Blur kernel estimation using the Radon transform. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 241–248, Colorado Springs, Colorado, June 2011. doi: 10.1109/CVPR.2011.5995479. (Cited on page 21.)

Peter Clifford. Markov random fields in statistics. In G. R. Grimmett and D. J. A. Welsh, editors, *Disorder in Physical Systems: A Volume in Honour of John Hammersley*, pages 19–32. Oxford University Press, 1990. ISBN 978-0198532156. (Cited on page 16.)

Richard Courant. Variational methods for the solution of problems of equilibrium and vibrations. *Bulletin of the American Mathematical Society*, 49:1–23, 1943. doi: 10.1090/S0002-9904-1943-07818-4. (Cited on page 15.)

Florent Couzinié-Devy, Jian Sun, Karteek Alahari, and Jean Ponce. Learning to estimate and remove non-uniform image blur. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1075–1082, Portland, Oregon, June 2013. doi: 10.1109/CVPR.2013.143. (Cited on pages 23, 79, 80, and 114.)

Shengyang Dai and Ying Wu. Removing partial blur in a single image. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 2544–2551, Miami, Florida, June 2009. doi: 10.1109/CVPR.2009.5206625. (Cited on pages 22, 78, 80, 85, 87, and 114.)

John G. Daugman. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *Journal of the Optical Society of America A*, 2 (7):1160–1169, July 1985. doi: 10.1364/JOSAA.2.001160. (Cited on page 16.)

A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B, Statistical Methodology*, 39(1):1–38, 1977. URL `http://www.jstor.org/stable/2984875`. (Cited on page x.)

Lawrence C. Evans. *Partial Differential Equations*, volume 19 of *Graduate Studies in Mathematics*. American Mathematical Society, 2nd edition, 2010. ISBN 978-0821849743. (Cited on pages 14 and 37.)

Mark Everingham, Luc Van Gool, Christopher K. I. Williams, John Winn, and Andrew Zisserman. The PASCAL visual object classes (VOC) challenge. *International Journal of Computer Vision*, 88(2):303–338, June 2010. doi: 10.1007/s11263-009-0275-4. (Cited on page 101.)

Mark Everingham, S. M. Ali Eslami, Luc Van Gool, Christopher K. I. Williams, John Winn, and Andrew Zisserman. The PASCAL visual object classes challenge: A retrospective. *International Journal of Computer Vision*, 111(1):98–136, January 2015. doi: 10.1007/s11263-014-0733-5. (Cited on pages xii and 101.)

Pedro F. Felzenszwalb and Daniel P. Huttenlocher. Efficient belief propagation for early vision. *International Journal of Computer Vision*, 1(70):41–54, October 2006. doi: 10.1007/s11263-006-7899-4. (Cited on pages 15 and 65.)

Rob Fergus, Barun Singh, Aaron Hertzmann, Sam T. Roweis, and William T. Freeman. Removing camera shake from a single photograph. *ACM Transactions on Graphics*, 25(3):787–794, July 2006. doi: 10.1145/1141911.1141956. (Cited on pages 18, 20, 49, 55, 56, 65, 72, 79, 94, 95, 101, and 103.)

Flickr. Camera finder. `http://www.flickr.com/cameras/`, 2017. Last accessed 4/2017. (Cited on page 1.)

Alessandro Foi, Mejdi Trimeche, Vladimir Katkovnik, and Karen Egiazarian. Practical Poissonian-Gaussian noise modeling and fitting for single-image raw-data. *IEEE Transactions on Image Processing*, 17(10):1737–1754, October 2008. doi: 10.1109/TIP.2008.2001399. (Cited on page 97.)

D. Gabor. Theory of communication. Part 1: The analysis of information. *Journal of the Institution of Electrical Engineers - Part III: Radio and Communication Engineering*, 93(26):429–441, 1946. doi: 10.1049/ji-3-2.1946.0074. (Cited on page 16.)

Qi Gao and Stefan Roth. How well do filter-based MRFs model natural images? In Axel Pinz, Thomas Pock, Horst Bischof, and Franz Leberl, editors, *Pattern Recognition, Proceedings of the joint 34th DAGM and 36th OAGM Symposium*, volume 7476 of *Lecture Notes in Computer Science*, pages 62–72. Springer, 2012. doi: 10.1007/978-3-642-32717-9_7. (Cited on pages 7, 17, and 102.)

Davi Geiger and Frederico Girosi. Parallel and deterministic algorithms from MRF's: Surface reconstruction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(5):401–412, May 1991. doi: 10.1109/34.134040. (Cited on pages 17, 64, and 65.)

Arnaud Gelas, Olivier Bernard, Denis Friboulet, and Rémy Prost. Compactly supported radial basis functions based collocation method for level-set evolution in image segmentation. *IEEE Transactions on Image Processing*, 16(7):1873–1887, July 2007. doi: 10.1109/TIP.2007.898969. (Cited on pages 15 and 30.)

Alan E. Gelfand and Adrian F. M. Smith. Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85(410):398–409, June 1990. doi: 10.2307/2289776. (Cited on page 16.)

I. M. Gelfand and S. V. Fomin. *Calculus of Variations*. Dover Books on Mathematics. Dover Publications, 2003. ISBN 978-0486414485. (Cited on page 14.)

D. Geman and Chengda Yang. Nonlinear image recovery with half-quadratic regularization. *IEEE Transactions on Image Processing*, 4 (7):932–946, July 1995. doi: 10.1109/83.392335. (Cited on pages 19 and 49.)

Donald Geman and George Reynolds. Constrained restoration and the recovery of discontinuities. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(3):367–383, March 1992. doi: 10.1109/34.120331. (Cited on page 19.)

Stuart Geman and Donald Geman. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:721–741, November 1984. doi: 10.1109/TPAMI.1984.4767596. (Cited on pages 7, 16, 28, 30, and 31.)

Timo Gissibl, Simon Thiele, Alois Herkommer, and Harald Giessen. Two-photon direct laser writing of ultracompact multi-lens objectives. *Nature Photonics*, 10(8):554–560, August 2016. doi: 10.1038/nphoton.2016.121. (Cited on page 1.)

Markus Grabner, Thomas Pock, Tobias Gross, and Bernhard Kainz. Automatic differentiation for GPU-accelerated 2D/3D registration. In Christian H. Bischof, H. Martin Bücker, Paul Hovland, Uwe Naumann, and Jean Utke, editors, *Advances in Automatic Differentiation*, volume 64 of *Lecture Notes in Computational Science and Engineering*, pages 259–269. Springer, 2008. doi: 10.1007/978-3-540-68942-3_23. (Cited on page 15.)

I. S. Gradshteyn and I. M. Ryzhik. *Table of Integrals, Series, and Products*. Academic Press, eighth edition, 2014. ISBN 978-0123849335. (Cited on pages 39 and 40.)

Christian Grossmann, Hans-Görg Roos, and Martin Stynes. *Numerical Treatment of Partial Differential Equations*. Universitext. Springer-Verlag Berlin Heidelberg, 2007. doi: 10.1007/978-3-540-71584-9. (Cited on page 14.)

Ankit Gupta, Neel Joshi, C. Lawrence Zitnick, Michael Cohen, and Brian Curless. Single image deblurring using motion density functions. In K. Daniilidis, P. Maragos, and N. Paragios, editors, *Proceedings of the 11th European Conference on Computer Vision*, Lecture

Notes in Computer Science, pages 171–184. Springer, 2010. doi: 10.1007/978-3-642-15549-9_13. (Cited on pages 5 and 21.)

Maya R. Gupta and Yihua Chen. Theory and use of the EM algorithm. *Foundations and Trends in Signal Processing*, 4(3):223–296, April 2011. doi: 10.1561/2000000034. (Cited on page 84.)

Alexandre Hammer, Jean Dumoulin, Benoi Vozel, and Kacem Chehdi. Deblurring of UAV aerial images for civil structures inspections using Mumford-Shah / total variation regularisation. In *5th International Symposium on Image and Signal Processing and Analysis*, Istanbul, Turkey, September 2007. doi: 10.1109/ISPA.2007.4383702. (Cited on page 2.)

J. M. Hammersley and P. Clifford. Markov fields on finite graphs and lattices. Not formally published, 1971. (Cited on page 16.)

Stefan Harmeling, Michael Hirsch, Suvrit Sra, and Bernhard Schölkopf. Online blind deconvolution for astronomical imaging. In *Proceedings of the IEEE International Conference on Computational Photography*, Pittsburgh, PA, April 2009. doi: 10.1109/ICCPHOT.2009.5559014. (Cited on page 3.)

Stefan Harmeling, Michael Hirsch, and Bernhard Schölkopf. Space-variant single-image blind deconvolution for removing camera shake. In J. Lafferty, C. K. I. Williams, J. Shawe-Taylor, R. S. Zemel, and A. Culotta, editors, *Advances in Neural Information Processing Systems*, volume 23, pages 829–837, 2010. (Cited on page 21.)

Magnus R. Hestenes and Eduard Stiefel. Methods of conjugate gradients for solving linear systems. *Journal of Research of the National Bureau of Standards*, 49(6):409–436, December 1952. doi: 10.6028/jres.049.044. (Cited on page x.)

Geoffrey E. Hinton. Products of experts. In *Proceedings of the Ninth International Conference on Artificial Neural Networks*, volume 1, pages 1–6, Edinburgh, UK, September 1999. doi: 10.1049/cp:19991075. (Cited on pages xi and 16.)

Geoffrey E. Hinton. Training products of experts by minimizing contrastive divergence. *Neural Computation*, 14(8):1771–1800, August 2002. doi: 10.1162/089976602760128018. (Cited on page 16.)

Jean-Baptiste Hiriart-Urruty and Claude Lemaréchal. *Convex Analysis and Minimization Algorithms I: Fundamentals*. Grundlehren der mathematischen Wissenschaften. Springer Berlin Heidelberg, 1993a. doi: 10.1007/978-3-662-02796-7. (Cited on pages 38 and 117.)

Jean-Baptiste Hiriart-Urruty and Claude Lemaréchal. *Convex Analysis and Minimization Algorithms II: Advanced Theory and Bundle Methods*. Grundlehren der mathematischen Wissenschaften. Springer

Berlin Heidelberg, 1993b. doi: 10.1007/978-3-662-06409-2. (Cited on pages 38 and 117.)

Michael Hirsch, Suvrit Sra, Bernhard Schölkopf, and Stefan Harmeling. Efficient filter flow for space-variant multiframe blind deconvolution. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 607–614, San Francisco, California, June 2010. doi: 10.1109/CVPR.2010.5540158. (Cited on pages x, 3, and 21.)

Michael Hirsch, Christian J. Schuler, Stefan Harmeling, and Bernhard Schölkopf. Fast removal of non-uniform camera shake. In *Proceedings of the Thirteenth IEEE International Conference on Computer Vision*, pages 463–470, Barcelona, Spain, November 2011. doi: 10.1109/ICCV.2011.6126276. (Cited on pages 21, 94, 102, 103, 104, and 109.)

Gerard J. Holzmann. *Beyond Photography: The Digital Darkroom*. Prentice Hall, March 1988. ISBN 978-0130744104. (Cited on page 4.)

Berthold K. P. Horn. *Robot Vision*. MIT Press, March 1986. ISBN 978-0262081597. (Cited on page 2.)

Roarke Horstmeyer. Camera motion tracking for deblurring and identification. Technical report, MIT Media Lab, 2010. (Cited on pages 22 and 116.)

Zhe Hu, Lu Yuan, Stephen Lin, and Ming-Hsuan Yang. Image deblurring using smartphone inertial sensors. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1855–1864, Las Vegas, Nevada, June 2016. doi: 10.1109/CVPR.2016.205. (Cited on page 22.)

Jinggang Huang. *Statistics of Natural Images and Models*. Ph.D. dissertation, Brown University, Division of Applied Mathematics, Providence, Rhode Island, May 2000. (Cited on page 7.)

Jinggang Huang and David Mumford. Statistics of natural images and models. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, pages 1541–1547, Fort Collins, Colorado, June 1999. doi: 10.1109/CVPR.1999.786990. (Cited on page 6.)

Hiroshi Ishikawa. Higher-order clique reduction in binary graph cut. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 2993–3000, Miami, Florida, June 2009. doi: 10.1109/CVPRW.2009.5206689. (Cited on pages 64, 69, 70, and 113.)

Ernst Ising. Beitrag zur Theorie des Ferromagnetismus. *Zeitschrift für Physik*, 31(1):253–258, February 1925. doi: 10.1007/BF02980577. (Cited on page 16.)

Tommi S. Jaakkola. Tutorial on variational approximation methods. In Manfred Opper and David Saad, editors, *Advanced Mean Field Methods: Theory and Practice*, Neural Information Processing, chapter 10. MIT Press, June 2001. ISBN 978-0262150545. (Cited on page 17.)

Viren Jain and H. Sebastian Seung. Natural image denoising with convolutional networks. In D. Koller, D. Schuurmans, Y. Bengio, and L. Bottou, editors, *Advances in Neural Information Processing Systems*, volume 21, pages 769–776, 2009. (Cited on page 43.)

Jeremy Jancsary, Sebastian Nowozin, Toby Sharp, and Carsten Rother. Regression tree fields — An efficient, non-parametric approach to image labeling problems. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 2376–2383, Providence, Rhode Island, June 2012a. doi: 10.1109/CVPR.2012.6247950. (Cited on pages xi, 8, 10, 24, 96, 98, and 114.)

Jeremy Jancsary, Sebastian Nowozin, and Carsten Rother. Loss-specific training of non-parametric image restoration models: A new state of the art. In A. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato, and C. Schmid, editors, *Proceedings of the 12th European Conference on Computer Vision*, volume 7578 of *Lecture Notes in Computer Science*, pages 112–125. Springer, 2012b. doi: 10.1007/978-3-642-33786-4_9. (Cited on pages 8, 10, 24, 96, and 98.)

E. T. Jaynes. Information theory and statistical mechanics. *Physical Review*, 106(4):620–630, May 1957. doi: 10.1103/PhysRev.106.620. (Cited on page 16.)

Claus S. Jensen, Uffe Kjærulff, and Augustine Kong. Blocking Gibbs sampling in very large probabilistic expert systems. *International Journal of Human-Computer Studies*, 42(6):647–666, June 1995. doi: 10.1006/ijhc.1995.1029. (Cited on pages 7 and 16.)

Claus Skaanning Jensen and Augustine Kong. Blocking Gibbs sampling for linkage analysis in large pedigrees with many loops. *The American Journal of Human Genetics*, 65(3):885–901, September 1999. doi: 10.1086/302524. (Cited on page 16.)

Jiaya Jia. Single image motion deblurring using transparency. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Minneapolis, Minnesota, June 2007. doi: 10.1109/CVPR.2007.383029. (Cited on pages 22, 78, and 80.)

Jiaya Jia. Robust deblurring software package v3.1. http://www.cse.cuhk.edu.hk/leojia/deblurring.htm, December 2013. Last accessed 4/2018. (Cited on page 94.)

Ming Jiang, Ge Wang, Margaret W. Skinner, Jay T. Rubinstein, and Michael W. Vannier. Blind deblurring of spiral CT images. *IEEE Transactions on Medical Imaging*, 22(7):837–845, July 2003. doi: 10.1109/TMI.2003.815075. (Cited on page 3.)

Hailin Jin, Paolo Favaro, and Roberto Cipolla. Visual tracking in the presence of motion blur. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, pages 18–25, San Diego, California, June 2005. doi: 10.1109/CVPR.2005.372. (Cited on page 2.)

Gunnar Johansson. Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics*, 14(2):201–211, June 1973. doi: 10.3758/BF03212378. (Cited on page 116.)

Michael I. Jordan, Zoubin Ghahramani, Tommi S. Jaakola, and Lawrence K. Saul. An introduction to variational methods for graphical models. *Machine Learning*, 37(2):183–233, November 1999. doi: 10.1023/A:1007665907178. (Cited on pages 7 and 17.)

Neel Joshi, Richard Szeliski, and David J. Kriegman. PSF estimation using sharp edge prediction. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Anchorage, Alaska, June 2008. doi: 10.1109/CVPR.2008.4587834. (Cited on pages 20, 49, and 95.)

Neel Joshi, Sing Bing Kang, C. Lawrence Zitnick, and Richard Szeliski. Image deblurring using inertial measurement sensors. *ACM Transactions on Graphics*, 29(4), July 2010. doi: 10.1145/1778765.1778767. (Cited on pages 21, 47, 49, 94, and 95.)

Byung Jun Kang and Kang Ryoung Park. Real-time image restoration for iris recognition systems. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 37(6):1555–1566, December 2007. doi: 10.1109/TSMCB.2007.907042. (Cited on page 2.)

Tal Kenig, Zvi Kam, and Arie Feuer. Blind image deconvolution using machine learning for three-dimensional microscopy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(12):2191–2204, December 2010. doi: 10.1109/TPAMI.2010.45. (Cited on page 3.)

Eamonn Keogh and Abdullah Mueen. Curse of dimensionality. In Claude Sammut and Geoffrey I. Webb, editors, *Encyclopedia of Machine Learning*, pages 257–258. Springer, 2010. doi: 10.1007/978-0-387-30164-8_192. (Cited on page 7.)

Margret Keuper, Thorsten Schmidt, Maja Temerinac-Ott, Jan Padeken, Patrick Heun, Olaf Ronneberger, and Thomas Brox. Blind deconvolution of widefield fluorescence microscopic data by regularization of the optical transfer function (OTF). In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 2179–2186, Portland, Oregon, June 2013. doi: 10.1109/CVPR.2013.283. (Cited on page 3.)

Tae Hyun Kim and Kyoung Mu Lee. Segmentation-free dynamic scene deblurring. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 2766–2773, Columbus, Ohio, June 2014. doi: 10.1109/CVPR.2014.348. (Cited on page 21.)

Tae Hyun Kim, Byeongjoo Ahn, and Kyoung Mu Lee. Dynamic scene deblurring. In *Proceedings of the Fourteenth IEEE International Conference on Computer Vision*, pages 3160–3167, Sydney, Australia, December 2013. doi: 10.1109/ICCV.2013.392. (Cited on pages 23 and 80.)

S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. Optimization by simulated annealing. *Science*, 220(4598):671–680, May 1983. doi: 10.1126/science.220.4598.671. (Cited on page 17.)

Rolf Köhler, Michael Hirsch, Betty Mohler, Bernhard Schölkopf, and Stefan Harmeling. Recording and playback of camera shake: Benchmarking blind deconvolution with a real-world database. In A. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato, and C. Schmid, editors, *Proceedings of the 12th European Conference on Computer Vision*, volume 7578 of *Lecture Notes in Computer Science*, pages 27–40. Springer, 2012. doi: 10.1007/978-3-642-33786-4_3. (Cited on pages 5, 20, 87, 95, 96, 102, 103, 104, and 105.)

Rolf Köhler, Michael Hirsch, Bernhard Schölkopf, and Stefan Harmeling. Improving alpha matting and motion blurred foreground estimation. In *Proceedings of the IEEE International Conference on Image Processing*, pages 3446–3450, Melbourne, Australia, September 2013. doi: 10.1109/ICIP.2013.6738711. (Cited on page 85.)

Vladmimir Kolmogorov and Ramin Zabih. What energy functions can be minimized via graph cuts? *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(2):147–159, February 2004. doi: 10.1109/TPAMI.2004.1262177. (Cited on page 15.)

Philipp Krähenbühl and Vladlen Koltun. Efficient inference in fully connected CRFs with Gaussian edge potentials. In J. Shawe-Taylor, R. S. Zemel, P. L. Bartlett, F. C. N. Pereira, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 24, pages 109–117, 2011. (Cited on pages 18 and 65.)

Kai Krajsek and Rudolf Mester. A maximum likelihood estimator for choosing the regularization parameters in global optical flow methods. In *Proceedings of the IEEE International Conference on Image Processing*, pages 1081–1084, Atlanta, Georgia, October 2006. doi: 10.1109/ICIP.2006.312743. (Cited on page 50.)

Dilip Krishnan and Rob Fergus. Fast image deconvolution using hyper-Laplacian priors. In Y. Bengio, D. Schuurmans, J. Lafferty, C. K. I. Williams, and A. Culotta, editors, *Advances in Neural Information Processing Systems*, volume 22, pages 1033–1041, 2009. (Cited on pages 19, 48, 49, 50, 57, 58, 59, 60, 61, 71, 72, 94, and 113.)

Dilip Krishnan, Terence Tay, and Rob Fergus. Blind deconvolution using a normalized sparsity measure. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 233–240, Colorado Springs, Colorado, June 2011. doi: 10.1109/CVPR.2011.5995521. (Cited on pages 20, 94, 95, 102, 103, 104, and 108.)

Dilip Krishnan, Raanan Fattal, and Richard Szeliski. Efficient preconditioning of Laplacian matrices for computer graphics. *ACM Transactions on Graphics*, 32(4):142:1–142:14, July 2013. doi: 10.1145/2461912.2461992. (Cited on page 14.)

S. R. Kulkarni, S. K. Mitter, T. J. Richardson, and J. N. Tsitsiklis. Local versus nonlocal computation of length of digitized curves. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(7):711–718, July 1994. doi: 10.1109/34.297951. (Cited on page 33.)

S. Kullback and R. A. Leibler. On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86, March 1951. URL `http://www.jstor.org/stable/2236703`. (Cited on page xi.)

M. Pawan Kumar, Vladimir Kolmogorov, and Philip H. S. Torr. An analysis of convex relaxations for MAP estimation of discrete MRFs. *Journal of Machine Learning Research*, 10:71–106, January 2009. (Cited on pages 29 and 31.)

Deepa Kundur and Dimitrios Hatzinakos. Blind image deconvolution. *IEEE Signal Processing Magazine*, 13(3):43–64, May 1996a. doi: 10.1109/79.489268. (Cited on pages 19 and 79.)

Deepa Kundur and Dimitrios Hatzinakos. Blind image deconvolution revisited. *IEEE Signal Processing Magazine*, 13(6):61–63, November 1996b. doi: 10.1109/79.543976. (Cited on page 19.)

Wei-Sheng Lai, Jia-Bin Huang, Zhe Hu, Narendra Ahuja, and Ming-Hsuan Yang. A comparative study for single image blind deblurring. In *Proceedings of the IEEE Computer Society Conference on*

Computer Vision and Pattern Recognition, pages 1701–1709, Las Vegas, Nevada, June 2016. doi: 10.1109/CVPR.2016.188. (Cited on page 20.)

Xiangyang Lan, Stefan Roth, Daniel P. Huttenlocher, and Michael J. Black. Efficient belief propagation with learned higher-order Markov random fields. In A. Leonardis, H. Bischof, and A. Pinz, editors, *Proceedings of the Ninth European Conference on Computer Vision*, volume 3952 of *Lecture Notes in Computer Science*, pages 269–282. Springer, 2006. doi: 10.1007/11744047_21. (Cited on pages 65, 69, and 70.)

Cornelius Lanczos. *Linear Differential Operators*. Martino Fine Books, 2012. ISBN 978-1614273028. (Cited on page 4.)

Ruxandra Lasowski, Art Tevs, Michael Wand, and Hans-Peter Seidel. Wavelet belief propagation for large scale inference problems. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1921–1928, Colorado Springs, Colorado, June 2011. doi: 10.1109/CVPR.2011.5995489. (Cited on pages 64 and 65.)

Wilhelm Lenz. Beitrag zum Verständnis der magnetischen Erscheinungen in festen Körpern. *Physikalische Zeitschrift*, 21:613–615, 1920. (Cited on page 16.)

Effi Levi. Using natural image priors – Maximizing or sampling? Master's thesis, The Hebrew University of Jerusalem, 2009. (Cited on pages 8, 17, 51, 53, and 58.)

A. Levin, D. Lischinski, and Y. Weiss. A closed form solution to natural image matting. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, pages 61–68, New York, New York, June 2006. doi: 10.1109/CVPR.2006.18. (Cited on page 22.)

Anat Levin. Blind motion deblurring using image statistics. In B. Schölkopf, J. Platt, and T. Hofmann, editors, *Advances in Neural Information Processing Systems*, volume 19, pages 841–848, 2007. (Cited on pages 6, 9, 23, 79, and 80.)

Anat Levin and Yair Weiss. User assisted separation of reflections from a single image using a sparsity prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(9):1647–1654, September 2007. doi: 10.1109/TPAMI.2007.1106. (Cited on pages 69, 73, and 74.)

Anat Levin, Rob Fergus, Frédo Durand, and William T. Freeman. Image and depth from a conventional camera with a coded aperture. *ACM Transactions on Graphics*, 26(3):70:1–70:9, July 2007. doi:

10.1145/1276377.1276464. (Cited on pages 19, 22, 48, 49, 50, 57, 58, 59, 60, 61, 71, 72, 80, 85, and 113.)

Anat Levin, Dani Lischinski, and Yair Weiss. A closed-form solution to natural image matting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(2):228–242, February 2008a. doi: 10.1109/TPAMI.2007.1177. (Cited on page 22.)

Anat Levin, Peter Sand, Taeg Sang Cho, Frédo Durand, and William T. Freeman. Motion-invariant photography. *ACM Transactions on Graphics*, 27(3):71:1–71:9, August 2008b. doi: 10.1145/1360612.1360670. (Cited on pages 22, 78, and 80.)

Anat Levin, Yair Weiss, Fredo Durand, and William T. Freeman. Understanding and evaluating blind deconvolution algorithms. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1964–1971, Miami, Florida, June 2009. doi: 10.1109/CVPRW.2009.5206815. (Cited on pages 5, 20, 47, 48, 49, 56, 58, 61, 95, 96, 107, 108, and 114.)

Anat Levin, Yair Weiss, Fredo Durand, and William T. Freeman. Efficient marginal likelihood optimization in blind deconvolution. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 2657–2664, Colorado Springs, Colorado, June 2011. doi: 10.1109/CVPR.2011.5995308. (Cited on pages 18, 20, 64, 65, 67, 68, 70, 71, 72, 79, 80, 82, 83, 84, 85, 86, 87, 89, 94, 95, 101, 103, 106, and 108.)

Stan Z. Li. *Markov Random Field Modeling in Image Analysis*. Advances in Pattern Recognition. Springer London, third edition, 2009. doi: 10.1007/978-1-84800-279-1. (Cited on page 15.)

Zhoubo Li, Lifeng Yu, Joshua D. Trzasko, David S. Lake, Daniel J. Blezek, Joel G. Fletcher, Cynthia H. McCollough, and Armando Manduca. Adaptive nonlocal means filtering based on local noise level for CT denoising. *Medical Physics*, 41(1), January 2014. doi: 10.1118/1.4851635. (Cited on page 3.)

L. B. Lucy. An iterative technique for the rectification of observed distributions. *The Astronomical Journal*, 79(6):745–754, June 1974. doi: 10.1086/111605. (Cited on pages 19, 49, 57, 58, 60, 61, 71, and 72.)

Siwei Lyu and Eero P. Simoncelli. Statistical modeling of images with fields of Gaussian scale mixtures. In B. Schölkopf, J. Platt, and T. Hofmann, editors, *Advances in Neural Information Processing Systems*, volume 19, pages 945–952, 2007. (Cited on page 43.)

John MacCormick and Andrew Fitzgibbon. Curvature regularization for resolution-independent images. In A. Heyden, F. Kahl, C. Olsson, M. Oskarsson, and X.-C. Tai, editors, *Proceedings of the 9th International Conference on Energy Minimization Methods in Computer*

Vision and Pattern Recognition, volume 8081 of *Lecture Notes in Computer Science*, pages 165–179, 2013. doi: 10.1007/978-3-642-40395-8_13. (Cited on page 15.)

Long Mai and Feng Liu. Kernel fusion for better image deblurring. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 371–380, Boston, Massachusetts, June 2015. doi: 10.1109/CVPR.2015.7298634. (Cited on pages 20, 24, and 116.)

Ju Man and Bir Bhanu. Individual recognition using gait energy image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(2):316–322, February 2006. doi: 10.1109/TPAMI.2006.38. (Cited on page 116.)

José V. Manjón, José Carbonell-Caballero, Juan J. Lull, Gracián García-Martí, Luís Martí-Bonmatí, and Montserrat Robles. MRI denoising using non-local means. *Medical Image Analysis*, 12(4):514–523, August 2008. doi: 10.1016/j.media.2008.02.004. (Cited on page 3.)

Manuel Martinello and Paolo Favaro. Fragmented aperture imaging for motion and defocus deblurring. In *Proceedings of the IEEE International Conference on Image Processing*, pages 3413–3416, Brussels, Belgium, September 2011. doi: 10.1109/ICIP.2011.6116444. (Cited on pages 22, 78, 80, 88, 90, and 114.)

Jerry M. Mendel. *Optimal Seismic Deconvolution: An Estimation Based Approach*. Academic Press, 1983. ISBN 978-1483243450. (Cited on page 19.)

Tomer Michaeli and Michal Irani. Nonparametric blind super-resolution. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 945–952, Portland, Oregon, June 2013. doi: 10.1109/ICCV.2013.121. (Cited on page 4.)

Tomer Michaeli and Michal Irani. Blind deblurring using internal patch recurrence. In D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, editors, *Proceedings of the 13th European Conference on Computer Vision*, Lecture Notes in Computer Science, pages 783–798. Springer, 2014. doi: 10.1007/978-3-319-10578-9_51. (Cited on page 95.)

Franck Michel. How many public photos are uploaded to Flickr every day, month, year? `https://www.flickr.com/photos/franckmichel/6855169886`, January 2017. Last accessed 4/2017. (Cited on page 1.)

Thomas Minka. Divergence measures and message passing. Technical Report MSR-TR-2005-173, Microsoft Research, Cambridge, UK, 2005. (Cited on pages 17, 66, 69, 82, and 114.)

Thomas P. Minka. Expectation propagation for approximate Bayesian inference. In *Proceedings of the Sixteenth Conference on Uncertainty in Artificial Intelligence*, pages 362–369, Seattle, Washington, August 2001. (Cited on pages x and 115.)

James Miskin and David J. C. MacKay. Ensemble learning for blind image separation and deconvolution. In Mark Girolami, editor, *Advances in Independent Component Analysis*, Perspectives in Neural Computing, pages 123–141. Springer London, 2000. doi: 10.1007/978-1-4471-0443-8. (Cited on pages 8, 17, 20, 49, 65, and 79.)

Bryan S. Morse, Weiming Liu, Terry S. Yoo, and Kalpathi Subramanian. Active contours using a constraint-based implicit representation. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, pages 285–292, San Diego, California, June 2005. doi: 10.1109/CVPR.2005.59. (Cited on pages 15 and 30.)

Patrick Naylor and Nikolay D. Gaubitch, editors. *Speech Dereverberation*. Signals and Communication Technology. Springer London, 2010. doi: 10.1007/978-1-84996-056-4. (Cited on page 19.)

Radford M. Neal. Probabilistic inference using Markov chain Monte Carlo methods. Technical Report CRG-TR-93-1, Department of Computer Science, University of Toronto, Ontario, Canada, September 1993. (Cited on page 16.)

Radford M. Neal and Geoffrey E. Hinton. A view of the EM algorithm that justifies incremental, sparse, and other variants. In Michael I. Jordan, editor, *Learning in Graphical Models*, Adaptive Computation and Machine Learning, pages 355–368. MIT Press, January 1999. ISBN 978-0262600323. (Cited on page 17.)

Masashi Nishiyama, Abdenour Hadid, Hidenori Takeshima, Jamie Shotton, Tatsuo Kozakaya, and Osamu Yamaguchi. Facial deblur inference using subspace analysis for recognition of blurred faces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33 (4):838–845, April 2011. doi: 10.1109/TPAMI.2010.203. (Cited on page 2.)

Sebastian Nowozin, Carsten Rother, Shai Bagon, Toby Sharp, Bangpeng Yao, and Pushmeet Kohli. Decision tree fields. In *Proceedings of the Thirteenth IEEE International Conference on Computer Vision*, pages 1668–1675, Barcelona, Spain, November 2011. doi: 10.1109/ICCV.2011.6126429. (Cited on page 8.)

Stanley Osher and Leonid I. Rudin. Feature-oriented image enhancement using shock filters. *SIAM Journal on Numerical Analysis*, 27(4): 919–940, August 1990. doi: 10.1137/0727053. (Cited on page 21.)

Christopher J. Pal, Jerod J. Weinman, Lam C. Tran, and Daniel Scharstein. On learning conditional random fields for stereo. *International Journal of Computer Vision*, 99(3):319–337, September 2012. doi: 10.1007/s11263-010-0385-z. (Cited on page 18.)

George Papandreou and Alan Yuille. Gaussian sampling by local perturbations. In J. Lafferty, C. K. I. Williams, J. Shawe-Taylor, R. S. Zemel, and A. Culotta, editors, *Advances in Neural Information Processing Systems*, volume 23, pages 1858–1866, 2010. (Cited on page 54.)

George Papandreou and Alan L. Yuille. Efficient variational inference in large-scale Bayesian compressed sensing. In *Proceedings of the Thirteenth IEEE International Conference on Computer Vision Workshops*, pages 1332–1339, Barcelona, Spain, November 2011. doi: 10.1109/ICCVW.2011.6130406. (Cited on page 115.)

Nils Papenberg, Andrés Bruhn, Thomas Brox, Stephan Didas, and Joachim Weickert. Highly accurate optic flow computation with theoretically justified warping. *International Journal of Computer Vision*, 67(2):141–158, April 2006. doi: 10.1007/s11263-005-3960-y. (Cited on page 29.)

Giorgio Parisi. *Statistical Field Theory*. Westview Press (Perseus), 1998. ISBN 978-0738200514. (Cited on page 17.)

Judea Pearl. Reverend Bayes on inference engines: A distributed hierarchical approach. In *Proceedings of the Second National Conference on Artificial Intelligence*, pages 133–136, Pittsburgh, Pennsylvania, August 1982. AAAI Press. (Cited on pages x and 15.)

Judea Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, 1988. ISBN 978-1558604797. (Cited on page 15.)

A.P. Pentland. Interpolation using wavelet bases. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(4):410–414, April 1994. doi: 10.1109/34.277594. (Cited on page 14.)

Daniele Perrone and Paolo Favaro. Total variation blind deconvolution: The devil is in the details. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 2909–2916, Columbus, Ohio, June 2014. doi: 10.1109/CVPR.2014.372. (Cited on pages 20 and 95.)

Daniele Perrone and Paolo Favaro. A clearer picture of total variation blind deconvolution. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(6):1041–1055, June 2016. doi: 10.1109/TPAMI.2015.2477819. (Cited on page 20.)

Pew Research Center. Mobile fact sheet. http://www.pewinternet.org/fact-sheet/mobile/, January 2017. Last accessed 4/2017. (Cited on page 1.)

Patrick Pletscher, Sebastian Nowozin, Pushmeet Kohli, and Carsten Rother. Putting MAP back on the map. In Rudolf Mester and Michael Felsberg, editors, *Pattern Recognition, Proceedings of the 33rd DAGM-Symposium*, volume 6835 of *Lecture Notes in Computer Science*, pages 111–121. Springer, 2011. doi: 10.1007/978-3-642-23123-0_12. (Cited on pages 63 and 69.)

Thomas Pock, Michael Pock, and Horst Bischof. Algorithmic differentiation: Application to variational problems in computer vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(7): 1180–1193, July 2007. doi: 10.1109/TPAMI.2007.1044. (Cited on pages 6, 14, 29, 33, and 38.)

Javier Portilla, Vasily Strela, Martin J. Wainwright, and Eero P. Simoncelli. Image denoising using scale mixtures of Gaussians in the wavelet domain. *IEEE Transactions on Image Processing*, 12(11): 1338–1351, November 2003. doi: 10.1109/TIP.2003.818640. (Cited on pages 42, 43, 44, 64, and 112.)

Travis Portz, Li Zhang, and Hongrui Jiang. Optical flow in the presence of spatially-varying motion blur. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1752–1759, Providence, Rhode Island, June 2012. doi: 10.1109/CVPR.2012.6247871. (Cited on page 87.)

Brian Potetz. Efficient belief propagation for vision using linear constraint nodes. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Minneapolis, Minnesota, June 2007. doi: 10.1109/CVPR.2007.383094. (Cited on pages 64, 65, 68, 69, and 70.)

Sathish Ramani, Philippe Thévenaz, and Michael Unser. Regularized interpolation for noisy data. In *Proceedings of the 4th IEEE International Symposium on Biomedical Imaging*, pages 612–615, Washington, DC, April 2007. doi: 10.1109/ISBI.2007.356926. (Cited on pages 30, 34, and 35.)

Ramesh Raskar, Amit Agrawal, and Jack Tumblin. Coded exposure photography: Motion deblurring using fluttered shutter. *ACM Transactions on Graphics*, 25(3):795–804, July 2006. doi: 10.1145/1141911.1141957. (Cited on pages 22, 78, and 80.)

William Hadley Richardson. Bayesian-based iterative method of image restoration. *Journal of the Optical Society of America*, 62(1):55–59, January 1972. doi: 10.1364/JOSA.62.000055. (Cited on pages 19, 49, 57, 58, 60, and 61.)

Walter Ritz. Über eine neue Methode zur Lösung gewisser Variationsprobleme der mathematischen Physik. *Journal für die reine und angewandte Mathematik (Crelle's Journal)*, 135:1–61, January 1909. doi: 10.1515/crll.1909.135.1. (Cited on pages 14 and 34.)

Fabiano Romeiro and Todd Zickler. Blind reflectometry. In K. Daniilidis, P. Maragos, and N. Paragios, editors, *Proceedings of the 11th European Conference on Computer Vision*, volume 6312 of *Lecture Notes in Computer Science*, pages 45–58. Springer, 2010. doi: 10.1007/978-3-642-15549-9_4. (Cited on pages 65 and 115.)

Stefan Roth and Michael J. Black. Fields of experts: A framework for learning image priors. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, pages 860–867, San Diego, California, June 2005. doi: 10.1109/CVPR.2005.160. (Cited on pages 7, 17, 42, and 43.)

Stefan Roth and Michael J. Black. Fields of experts. *International Journal of Computer Vision*, 82(2):205–229, April 2009. doi: 10.1007/s11263-008-0197-6. (Cited on pages xi, 17, 31, 32, 48, 51, 56, 57, 58, 59, 60, 61, 64, 73, 112, and 113.)

Carsten Rother, Vladimir Kolmogorov, and Andrew Blake. "GrabCut" — interactive foreground extraction using iterated graph cuts. *ACM Transactions on Graphics*, 23(3):309–314, August 2004. doi: 10.1145/1015706.1015720. (Cited on pages 15 and 23.)

Leonid I. Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1–4):259–268, November 1992. doi: 10.1016/0167-2789(92)90242-F. (Cited on pages xi, 14, 27, 28, 29, 33, 38, 69, 112, and 117.)

Håvard Rue and Leonhard Held. *Gaussian Markov Random Fields: Theory and Applications*. Monographs on Statistics & Applied Probability. Chapman & Hall/CRC, February 2005. ISBN 978-1584884323. (Cited on page 16.)

Yousef Saad. *Iterative Methods for Sparse Linear Systems*. Other Titles in Applied Mathematics. Society for Industrial and Applied Mathematics, second edition, 2003. doi: 10.1137/1.9780898718003. (Cited on page 14.)

Pinaki Sarder and Arye Nehorai. Deconvolution methods for 3-D fluorescence microscopy images. *IEEE Signal Processing Magazine*, 23(3):32–45, May 2006. doi: 10.1109/MSP.2006.1628876. (Cited on page 3.)

Hanno Scharr, Michael J. Black, and Horst W. Haussecker. Image statistics and anisotropic diffusion. In *Proceedings of the Ninth IEEE*

International Conference on Computer Vision, volume 2, pages 840–847, Nice, France, October 2003. doi: 10.1109/ICCV.2003.1238435. (Cited on page 28.)

Kevin Schelten and Stefan Roth. Connecting non-quadratic variational models and MRFs. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 2641–2648, Colorado Springs, Colorado, June 2011a. doi: 10.1109/CVPR.2011.5995498. (Cited on page 10.)

Kevin Schelten and Stefan Roth. Connecting non-quadratic variational models and MRFs. Technical Report TUD-CS-2011-0103, TU Darmstadt, Department of Computer Science, 2011b. (Cited on page 10.)

Kevin Schelten and Stefan Roth. Mean field for continuous high-order MRFs. In Axel Pinz, Thomas Pock, Horst Bischof, and Franz Leberl, editors, *Pattern Recognition, Proceedings of the joint 34th DAGM and 36th OAGM Symposium*, volume 7476 of *Lecture Notes in Computer Science*, pages 52–61. Springer, 2012. doi: 10.1007/978-3-642-32717-9_6. (Cited on pages 8 and 10.)

Kevin Schelten and Stefan Roth. Localized image blur removal through non-parametric kernel estimation. In *Proceedings of the International Conference on Pattern Recognition*, pages 702–707, Stockholm, Sweden, August 2014. doi: 10.1109/ICPR.2014.131. (Cited on pages 6, 10, and 24.)

Kevin Schelten, Sebastian Nowozin, Jeremy Jancsary, Carsten Rother, and Stefan Roth. Interleaved regression tree field cascades for blind image deconvolution. In *Proceedings of the IEEE Winter Conference on Applications of Computer Vision*, pages 494–501, Waikoloa Beach, Hawaii, January 2015. doi: 10.1109/WACV.2015.72. (Cited on pages 8, 11, and 25.)

Uwe Schmidt. Learning and evaluating Markov random fields for natural images. M.Sc. thesis, TU Darmstadt, Germany, February 2010. (Cited on pages 52 and 53.)

Uwe Schmidt. *Half-quadratic Inference and Learning for Natural Images*. Ph.D. dissertation, TU Darmstadt, Department of Computer Science, Darmstadt, Germany, December 2016. (Cited on pages 9, 10, and 54.)

Uwe Schmidt and Stefan Roth. Shrinkage fields for effective image restoration. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 2774–2781, Columbus, Ohio, June 2014. doi: 10.1109/CVPR.2014.349. (Cited on pages 24, 94, and 96.)

Uwe Schmidt, Qi Gao, and Stefan Roth. A generative perspective on MRFs in low-level vision. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1751–1758, San Francisco, California, June 2010. doi: 10.1109/CVPR.2010.5539844. (Cited on pages 7, 9, 17, 32, 48, 49, 51, 52, 53, 54, 56, 64, 66, 69, 70, and 112.)

Uwe Schmidt, Kevin Schelten, and Stefan Roth. Bayesian deblurring with integrated noise estimation. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 2625–2632, Colorado Springs, Colorado, June 2011. doi: 10.1109/CVPR.2011.5995653. (Cited on pages 7, 10, 19, and 94.)

Uwe Schmidt, Carsten Rother, Sebastian Nowozin, Jeremy Jancsary, and Stefan Roth. Discriminative non-blind deblurring. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 604–611, Portland, Oregon, June 2013. doi: 10.1109/CVPR.2013.84. (Cited on pages 8, 24, 94, 96, 98, 99, 101, 102, and 103.)

Uwe Schmidt, Jeremy Jancsary, Sebastian Nowozin, Stefan Roth, and Carsten Rother. Cascades of regression tree fields for image restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(4):677–689, April 2016. doi: 10.1109/TPAMI.2015.2441053. (Cited on pages 24, 98, 114, and 115.)

Christoph Schnörr, Rainer Sprengel, and Bernd Neumann. A variational approach to the design of early vision algorithms. In W. Kropatsch, R. Klette, F. Solina, and R. Albrecht, editors, *Theoretical Foundations of Computer Vision*, volume 11 of *Computing Supplement*, pages 149–165. Springer, 1996. doi: 10.1007/978-3-7091-6586-7_9. (Cited on page 27.)

Christian J. Schuler, Harold Christopher Burger, Stefan Harmeling, and Bernhard Schölkopf. A machine learning approach for non-blind image deconvolution. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1067–1074, Portland, Oregon, June 2013. doi: 10.1109/CVPR.2013.142. (Cited on pages 24, 94, and 96.)

Christian J. Schuler, Michael Hirsch, Stefan Harmeling, and Bernhard Schölkopf. Learning to deblur. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(7):1439–1451, July 2016. doi: 10.1109/TPAMI.2015.2481418. (Cited on pages 24 and 114.)

Matthias Seeger and Hannes Nickisch. Fast convergent algorithms for expectation propagation approximate Bayesian inference. In Geoffrey Gordon, David Dunson, and Miroslav Dudík, editors, *Proceedings of the Fourteenth International Workshop on Artificial Intelligence*

and Statistics, pages 652–660, Ft. Lauderdale, Florida, April 2011. (Cited on page 115.)

Qi Shan, Wei Xiong, and Jiaya Jia. Rotational motion deblurring of a rigid object from a single image. In *Proceedings of the Eleventh IEEE International Conference on Computer Vision*, Rio de Janeiro, Brazil, October 2007. doi: 10.1109/ICCV.2007.4408922. (Cited on page 22.)

Qi Shan, Jiaya Jia, and Aseem Agarwala. High-quality motion deblurring from a single image. *ACM Transactions on Graphics*, 27(3): 73:1–73:10, August 2008. doi: 10.1145/1360612.1360672. (Cited on pages 49 and 95.)

Jianping Shi, Li Xu, and Jiaya Jia. Discriminative blur detection features. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 2965–2972, Columbus, Ohio, June 2014. doi: 10.1109/CVPR.2014.379. (Cited on pages 23 and 86.)

Till Sieberth, Rene Wackrow, and Jim H. Chandler. Motion blur disturbs – the influence of motion-blurred images in photogrammetry. *The Photogrammetric Record*, 29(148):434–453, December 2014. doi: 10.1111/phor.12082. (Cited on page 2.)

Ondřej Šindelář and Filip Šroubek. Image deblurring in smartphone devices using built-in inertial measurement sensors. *Journal of Electronic Imaging*, 22(1), 2013. doi: 10.1117/1.JEI.22.1.011003. (Cited on page 22.)

Ondřej Šindelář, Filip Šroubek, and Peyman Milanfar. Space-variant image deblurring on smartphones using inertial sensors. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pages 191–192, Columbus, Ohio, June 2014a. doi: 10.1109/CVPRW.2014.34. (Cited on page 22.)

Ondřej Šindelář, Filip Šroubek, and Peyman Milanfar. A smartphone application for removing handshake blur and compensating rolling shutter. In *Proceedings of the IEEE International Conference on Image Processing*, pages 2160–2162, Paris, France, October 2014b. doi: 10.1109/ICIP.2014.7025433. (Cited on page 22.)

Mark Sollweck. Sampling-based Bayesian inference for optical flow. M.Sc. thesis, TU Darmstadt, Germany, December 2012. (Cited on page 7.)

J. Stoer and R. Bulirsch. *Introduction to Numerical Analysis*, volume 12 of *Texts in Applied Mathematics*. Springer New York, 2002. doi: 10.1007/978-0-387-21738-3. (Cited on page 41.)

Erik B. Sudderth, Alexander T. Ihler, William T. Freeman, and Alan S. Willsky. Nonparametric belief propagation. In *Proceedings of the*

IEEE Computer Society Conference on Computer Vision and Pattern Recognition, volume 1, pages 605–612, Madison, Wisconsin, June 2003. doi: 10.1109/CVPR.2003.1211409. (Cited on page 65.)

Jian Sun, Wenfei Cao, Zongben Xu, and Jean Ponce. Learning a convolutional neural network for non-uniform motion blur removal. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 769–777, Boston, Massachusetts, June 2015. doi: 10.1109/CVPR.2015.7298677. (Cited on pages 21 and 23.)

Libin Sun, Sunghyun Cho, Jue Wang, and James Hays. Edge-based blur kernel estimation using patch priors. In *Proceedings of the Fifth IEEE International Conference on Computational Photography*, Cambridge, MA, April 2013. doi: 10.1109/ICCPhot.2013.6528301. (Cited on pages 20, 21, and 95.)

Libin Sun, Sunghyun Cho, Jue Wang, and James Hays. Good image priors for non-blind deconvolution: Generic vs specific. In D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, editors, *Proceedings of the 13th European Conference on Computer Vision*, volume 8692 of *Lecture Notes in Computer Science*, pages 231–246. Springer, 2014. doi: 10.1007/978-3-319-10593-2_16. (Cited on page 95.)

Richard Szeliski. Estimating motion from sparse range data without correspondence. In *Proceedings of the Second IEEE International Conference on Computer Vision*, pages 207–215, Tampa, Florida, December 1988. doi: 10.1109/CCV.1988.589992. (Cited on page 18.)

Richard Szeliski. Bayesian modeling of uncertainty in low-level vision. *International Journal of Computer Vision*, 5(3):271–301, December 1990a. doi: 10.1007/BF00126502. (Cited on pages 9, 18, 28, 32, 33, 36, 37, and 112.)

Richard Szeliski. Fast surface interpolation using hierarchical basis functions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(6):513–528, June 1990b. doi: 10.1109/34.56188. (Cited on page 14.)

Richard Szeliski. Locally adapted hierarchical basis preconditioning. *ACM Transactions on Graphics*, 25(3):1135–1143, July 2006. doi: 10.1145/1179352.1142005. (Cited on page 14.)

Yu-Wing Tai, Hao Du, Michael S. Brown, and Stephen Lin. Image/video deblurring using a hybrid camera. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Anchorage, Alaska, June 2008. doi: 10.1109/CVPR.2008.4587507. (Cited on page 49.)

Yu-Wing Tai, Ping Tan, and Michael S. Brown. Richardson-Lucy deblurring for scenes under a projective motion path. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(8):1603–1618, August 2011. doi: 10.1109/TPAMI.2010.222. (Cited on pages 5 and 21.)

Dacheng Tao, Xuelong Li, Xindong Wu, and Stephen J. Maybank. General tensor discriminant analysis and gabor features for gait recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(10):1700–1715, October 2007. doi: 10.1109/TPAMI.2007.1096. (Cited on page 116.)

Marshall F. Tappen, Ce Liu, Edward H. Adelson, and William T. Freeman. Learning Gaussian conditional random fields for low-level vision. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Minneapolis, Minnesota, June 2007. doi: 10.1109/CVPR.2007.382979. (Cited on page 95.)

Yeh Whye Teh, Max Welling, Simon Osindero, and Geoffrey E. Hinton. Energy-based models for sparse overcomplete representations. *Journal of Machine Learning Research*, 4:1235–1260, December 2003. (Cited on page 16.)

Demetri Terzopoulos. Multilevel computational processes for visual surface reconstruction. *Computer Vision, Graphics, and Image Processing*, 24(1):52–96, October 1983. doi: 10.1016/0734-189X(83)90020-8. (Cited on pages 15 and 30.)

Demetri Terzopoulos. Multilevel reconstruction of visual surfaces: Variational principles and finite-element representations. In Azriel Rosenfeld, editor, *Multiresolution Image Processing and Analysis*, volume 12 of *Springer Series in Information Sciences*, pages 237–310. Springer Berlin Heidelberg, 1984. doi: 10.1007/978-3-642-51590-3_17. (Cited on page 15.)

Demetri Terzopoulos. Regularization of inverse visual problems involving discontinuities. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(4):413–424, July 1986. doi: 10.1109/TPAMI.1986.4767807. (Cited on pages 15, 18, 30, and 32.)

Demetri Terzopoulos. The computation of visible-surface representations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10(4):417–438, July 1988. doi: 10.1109/34.3908. (Cited on pages 15, 18, 30, and 34.)

A. N. Tikhonov and V. Y. Arsenin. *Solutions of Ill-Posed Problems*. V. H. Winston (Wiley), 1977. ISBN 978-0470991244. (Cited on page 13.)

C. Tomasi and R. Manduchi. Bilateral filtering for gray and color images. In *Proceedings of the Sixth IEEE International Conference on*

Computer Vision, pages 839–846, Bombay, India, January 1998. doi: 10.1109/ICCV.1998.710815. (Cited on page 21.)

Irina Topalova and Alexander Tzokev. Increasing the image recognition accuracy in machine vision systems with added noise due to technological issues. In *IEEE 26th Convention of Electrical and Electronics Engineers in Israel*, pages 000328–000332, Eliat, Israel, November 2010. doi: 10.1109/EEEI.2010.5662212. (Cited on page 2.)

Robert Nigel Tubbs. *Lucky Exposures: Diffraction limited astronomical imaging through the atmosphere*. Ph.D. dissertation, Cambridge University, Department of Physics, Cambridge, UK, September 2003. (Cited on page 3.)

Jack Tzeng, Chun-Chen Liu, and Truong Q. Nguyen. Contourlet domain multiband deblurring based on color correlation for fluid lens cameras. *IEEE Transactions on Image Processing*, 19(10):2659–2668, October 2010. doi: 10.1109/TIP.2010.2049237. (Cited on page 3.)

Vladimir N. Vapnik. *Statistical Learning Theory*. Adaptive and Learning Systems for Signal Processing, Communications, and Control. John Wiley & Sons, October 1998. ISBN 978-0471030034. (Cited on page 8.)

Vibhav Vineet, Jonathan Warrell, and Philip H. S. Torr. Filter-based mean-field inference for random fields with higher-order terms and product label-spaces. *International Journal of Computer Vision*, 110(3): 290–307, December 2014. doi: 10.1007/s11263-014-0708-6. (Cited on page 18.)

Fabio Viola, Roberto Cipolla, and Andrew Fitzgibbon. A unifying resolution-independent formulation for early vision. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 494–501, Providence, Rhode Island, June 2012. doi: 10.1109/CVPR.2012.6247713. (Cited on page 15.)

Andrey Vlasenko and Christoph Schnörr. Superresolution and denoising of 3D fluid flow estimates. In Gerhard Rigoll, editor, *Pattern Recognition, Proceedings of the 31st DAGM-Symposium*, volume 5748 of *Lecture Notes in Computer Science*, pages 482–491. Springer, 2009. doi: 10.1007/978-3-642-03798-6_49. (Cited on page 30.)

Martin J. Wainwright and Michael I. Jordan. Graphical models, exponential families, and variational inference. *Foundations and Trends in Machine Learning*, 1(1–2):1–305, 2008. doi: 10.1561/2200000001. (Cited on page 17.)

Martin J. Wainwright and Eero P. Simoncelli. Scale mixtures of Gaussians and the statistics of natural images. In S. A. Solla, T. K. Leen,

and K.-R. Müller, editors, *Advances in Neural Information Processing Systems*, volume 12, pages 855–861, 2000. (Cited on pages 51 and 66.)

Martin J. Wainwright, Tommi S. Jaakkola, and Alan S. Willsky. MAP estimation via agreement on trees: Message passing and linear programming. *IEEE Transactions on Information Theory*, 51(11):3697–3717, November 2005. doi: 10.1109/TIT.2005.856938. (Cited on pages 29 and 31.)

Jue Wang. Image and video deblurring. http://www.juew.org/blur/deblur.htm, July 2013. Last accessed 4/2017. (Cited on page 4.)

Liang Wang, Tieniu Tan, Huazhong Ning, and Weiming Hu. Silhouette analysis-based gait recognition for human identification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(12):1505–1518, December 2003. doi: 10.1109/TPAMI.2003.1251144. (Cited on page 116.)

Ruxin Wang and Dacheng Tao. Recent progress in image deblurring. *arXiv:1409.6838*, 2014. URL http://arxiv.org/abs/1409.6838. (Cited on pages 19, 79, and 94.)

Zhaozhong Wang, Feihu Qi, and Fugen Zhou. A discontinuous finite element method for image denoising. In Aurélio Campilho and Mohamed S. Kamel, editors, *Proceedings of the Third International Conference on Image Analysis and Recognition*, volume 4141 of *Lecture Notes in Computer Science*, pages 116–125. Springer, 2006. doi: 10.1007/11867586_11. (Cited on page 30.)

Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, April 2004. doi: 10.1109/TIP.2003.819861. (Cited on page xii.)

Allan G. Weber. The USC-SIPI image database: Version 6. Technical Report USC-SIPI-432, Signal and Image Processing Institute, Ming Hsieh Department of Electrical Engineering, University of Southern California, Los Angeles, California, February 2018. (Cited on page 45.)

Jerod J. Weinman, Lam Tran, and Christopher J. Pal. Efficiently learning random fields for stereo vision with sparse message passing. In *Proceedings of the Tenth European Conference on Computer Vision*, volume 5302 of *Lecture Notes in Computer Science*, pages 617–630. Springer, 2008. doi: 10.1007/978-3-540-88682-2_47. (Cited on page 18.)

Mark Weiser. The computer for the 21st century. *Scientific American*, 265(3):94–104, September 1991. doi: 10.1038/scientificamerican0991-94. (Cited on page 1.)

Yair Weiss and William T. Freeman. What makes a good model of natural images? In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Minneapolis, Minnesota, June 2007. doi: 10.1109/CVPR.2007.383092. (Cited on pages 7, 17, and 66.)

Max Welling, Geoffrey E. Hinton, and Simon Osindero. Learning sparse topographic representations with products of Student-t distributions. In S. Becker, S. Thrun, and K. Obermayer, editors, *Advances in Neural Information Processing Systems*, volume 15, pages 1359–1366, 2003. (Cited on page 16.)

Manuel Werlberger, Werner Trobin, Thomas Pock, Andreas Wedel, Daniel Cremers, and Horst Bischof. Anisotropic Huber-L1 optical flow. In *Proceedings of the British Machine Vision Conference*, London, UK, September 2009. (Cited on page 43.)

Oliver Whyte, Josef Sivic, Andrew Zisserman, and Jean Ponce. Non-uniform deblurring for shaken images. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 491–498, San Francisco, California, June 2010. doi: 10.1109/CVPR.2010.5540175. (Cited on pages 5, 21, 78, 79, 80, and 94.)

Oliver Whyte, Josef Sivic, and Andrew Zisserman. Deblurring shaken and partially saturated images. In *Proceedings of the Thirteenth IEEE International Conference on Computer Vision Workshops*, pages 745–752, Barcelona, Spain, November 2011. doi: 10.1109/ICCVW.2011.6130327. (Cited on pages 21, 102, 103, and 104.)

Norbert Wiener. *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*. MIT Press, March 1964. ISBN 978-0262730051. (Cited on pages 5 and 24.)

Ralph A. Wiggins. Minimum entropy deconvolution. *Geoexploration*, 16(1–2):21–35, April 1978. doi: 10.1016/0016-7142(78)90005-4. (Cited on page 19.)

John Winn and Christopher M. Bishop. Variational message passing. *Journal of Machine Learning Research*, 6:661–694, April 2005. (Cited on pages 7, 17, and 66.)

David Wipf and Haichao Zhang. Analysis of Bayesian blind deconvolution. In A. Heyden, F. Kahl, C. Olsson, M. Oskarsson, and X.-C. Tai, editors, *Proceedings of the 9th International Conference on Energy Minimization Methods in Computer Vision and Pattern Recognition*, volume 8081 of *Lecture Notes in Computer Science*, pages 40–53, 2013. doi: 10.1007/978-3-642-40395-8_4. (Cited on pages 79, 80, and 95.)

David Wipf and Haichao Zhang. Revisiting bayesian blind deconvolution. *Journal of Machine Learning Research*, 15:3775–3814, November 2014. (Cited on pages 8, 18, 20, and 115.)

Yi Wu, Haibin Ling, Jingyi Yu, Feng Li, Xue Mei, and Erkang Cheng. Blurred target tracking by blur-driven tracker. In *Proceedings of the Thirteenth IEEE International Conference on Computer Vision*, pages 1100–1107, Barcelona, Spain, November 2011. doi: 10.1109/ICCV.2011.6126357. (Cited on page 2.)

Lei Xiao, Jue Wang, Wolfgang Heidrich, and Michael Hirsch. Learning high-order filters for efficient blind deconvolution of document photographs. In B. Leibe, J. Matas, N. Sebe, and M. Welling, editors, *Proceedings of the 14th European Conference on Computer Vision*, volume 9907 of *Lecture Notes in Computer Science*, pages 734–749. Springer, 2016. doi: 10.1007/978-3-319-46487-9_45. (Cited on pages 25 and 114.)

Li Xu and Jiaya Jia. Two-phase kernel estimation for robust motion deblurring. In K. Daniilidis, P. Maragos, and N. Paragios, editors, *Proceedings of the 11th European Conference on Computer Vision*, volume 6311 of *Lecture Notes in Computer Science*, pages 157–170. Springer, 2010. doi: 10.1007/978-3-642-15549-9_12. (Cited on pages 5, 21, 78, 86, 94, 95, 99, 100, 102, 103, 104, 107, and 108.)

Li Xu, Shicheng Zheng, and Jiaya Jia. Unnatural L_0 sparse representation for natural image deblurring. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1107–1114, Portland, Oregon, June 2013. doi: 10.1109/CVPR.2013.147. (Cited on pages 6, 21, 94, 95, 101, and 103.)

Jonathan S. Yedidia, William T. Freeman, and Yair Weiss. Understanding belief propagation and its generalizations. In G. Lakemeyer and B. Nebel, editors, *Exploring Artificial Intelligence in the New Millennium*, chapter 8, pages 239–269. Morgan Kaufmann, 2003. ISBN 978-1558608115. (Cited on pages 15, 29, and 31.)

L. Yuan, J. Sun, L. Quan, and H.Y. Shum. Image deblurring with blurred/noisy image pairs. *ACM Transactions on Graphics*, 26(3):1:1–1:10, July 2007. doi: 10.1145/1276377.1276379. (Cited on page 49.)

Christopher Zach, Thomas Pock, and Horst Bischof. A duality based approach for realtime TV-L^1 optical flow. In Fred A. Hamprecht, Christoph Schnörr, and Bernd Jähne, editors, *Pattern Recognition, Proceedings of the 29th DAGM-Symposium*, volume 4713 of *Lecture Notes in Computer Science*, pages 214–223. Springer, 2007. doi: 10.1007/978-3-540-74936-3_22. (Cited on pages 6 and 14.)

Li Zhang and S. M. Seitz. Estimating optimal parameters for MRF stereo from a single image pair. *IEEE Transactions on Pattern Anal-*

ysis and Machine Intelligence, 29(2):331–342, February 2007. doi: 10.1109/TPAMI.2007.36. (Cited on page 50.)

Wei Zhang and Wai-Kuen Cham. Single image focus editing. In *Proceedings of the Twelfth IEEE International Conference on Computer Vision Workshops*, pages 1947–1954, Kyoto, Japan, September–October 2009. doi: 10.1109/ICCVW.2009.5457520. (Cited on page 86.)

Song Chun Zhu and David Mumford. Prior learning and Gibbs reaction-diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(11):1236–1250, November 1997. doi: 10.1109/34.632983. (Cited on pages 7 and 16.)

Song Chun Zhu, Ying Nian Wu, and David Mumford. Minimax entropy principle and its application to texture modeling. *Neural Computation*, 9(8):1627–1660, November 1997. doi: 10.1162/neco.1997.9.8.1627. (Cited on page 16.)

Song Chun Zhu, Yingnian Wu, and David Mumford. Filters, random fields and maximum entropy (FRAME): Towards a unified theory for texture modeling. *International Journal of Computer Vision*, 27 (2):107–126, March 1998. doi: 10.1023/A:1007925832420. (Cited on pages xi and 16.)

Xiang Zhu and Peyman Milanfar. Removing atmospheric turbulence via space-invariant deconvolution. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(1):157–170, January 2013. doi: 10.1109/TPAMI.2012.82. (Cited on page 3.)

Daniel Zoran and Yair Weiss. Scale invariance and noise in natural images. In *Proceedings of the Twelfth IEEE International Conference on Computer Vision*, pages 2209–2216, Kyoto, Japan, September–October 2009. doi: 10.1109/ICCV.2009.5459476. (Cited on pages 57, 58, 73, and 113.)

Daniel Zoran and Yair Weiss. From learning models of natural image patches to whole image restoration. In *Proceedings of the Thirteenth IEEE International Conference on Computer Vision*, pages 479–486, Barcelona, Spain, November 2011. doi: 10.1109/ICCV.2011.6126278. (Cited on pages x, 20, and 94.)

Daniel Zoran and Yair Weiss. Natural images, Gaussian mixtures and dead leaves. In P. Bartlett, F. C. N. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 25, pages 1736–1744, 2012. (Cited on page 20.)

Wangmeng Zuo, Dongwei Ren, Shuhang Gu, Liang Lin, and Lei Zhang. Discriminative learning of iteration-wise priors for blind deconvolution. In *Proceedings of the IEEE Computer Society Conference*

on Computer Vision and Pattern Recognition, pages 3232–3240, Boston, Massachusetts, June 2015. doi: 10.1109/CVPR.2015.7298943. (Cited on pages 24 and 114.)

PUBLICATIONS

Kevin Schelten, Sebastian Nowozin, Jeremy Jancsary, Carsten Rother, and Stefan Roth.
Interleaved regression tree field cascades for blind image deconvolution.
In *Proceedings of the IEEE Winter Conference on Applications of Computer Vision*, pages 494–501, Waikoloa Beach, Hawaii, January 2015.

Kevin Schelten and Stefan Roth.
Localized image blur removal through non-parametric kernel estimation.
In *Proceedings of the International Conference on Pattern Recognition*, pages 702–707, Stockholm, Sweden, August 2014.

Kevin Schelten and Stefan Roth.
Mean field for continuous high-order MRFs.
In *Pattern Recognition, Proceedings of the joint 34th DAGM and 36th OAGM Symposium*, volume 7476 of *Lecture Notes in Computer Science*. Springer, 2012.

Kevin Schelten and Stefan Roth.
Connecting non-quadratic variational models and MRFs.
In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 2641–2648, Colorado Springs, Colorado, June 2011a.

Kevin Schelten and Stefan Roth.
Connecting non-quadratic variational models and MRFs.
Technical Report TUD-CS-2011-0103, TU Darmstadt, Department of Computer Science, 2011b.

Uwe Schmidt, Kevin Schelten, and Stefan Roth.
Bayesian deblurring with integrated noise estimation.
In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 2625–2632, Colorado Springs, Colorado, June 2011.